UNIVERSITY of TASMANIA

Understanding the Molecular Epidemiology of *Mycobacterium tuberculosis* Infection from Whole-genome Analyses

by

Sanjay Gautam

College of Health and Medicine

Submitted in fulfilment of the requirements for the Doctor of Philosophy (PhD)

University of Tasmania

December 2018

Declaration of Originality

“This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by any other person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.”

_____

Sanjay Gautam

Statement of Authority of Access

_____

Sanjay Gautam

Statement Regarding Published Work Contained in Thesis

“The publishers of the papers comprising chapter 3, chapter 4 and chapter 5 hold the copyright for the respective content, and access to the material should be sought from the respective journals. The remaining non-published content of the thesis may be made available for loan and limited copying in accordance with copyright Act 1968.”

_____

Sanjay Gautam

Statement of Ethical Conduct

“The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government’s Office of the Gene Technology Regulator, The Tasmania Health and Medical Human Research Ethics Committee, and the rulings of the safety, Ethics and Institutional Biosafety Committees of the University.”

_____

Sanjay Gautam

Publications

‡Data presented in these papers/abstracts are discussed in the thesis. *Corresponding author.

Peer reviewed original research articles:

‡**Gautam SS**, Haug G, Cooley LA, Mac Aogáin M and O'Toole RF*. Potential cross-border translocation of multi-drug resistant tuberculosis in latent form to Australia. *Medical Journal of Australia* 2019, 210(5), doi: 10.5694/mja2.50044

‡**Gautam SS**, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, Globan M and O'Toole RF*. Molecular Epidemiology of Tuberculosis in Tasmania and Genomic Characterisation of its First Known Multi-Drug Resistant Case. *PLoS ONE* 2018; 13(2): e0192351, doi: 10.1371/journal.pone.0192351.

‡**Gautam SS,** Mac Aogáin M, Bower JE, Basu I and O'Toole RF*. Differential Carriage of Virulence-associated Loci in the New Zealand Rangipo Outbreak Strain of *Mycobacterium tuberculosis*. *Infectious Diseases* 2017; 49(9) 680-8, doi: 10.1080/23744235.2017.1330553.

‡**Gautam SS,** KC R, Leong KWC, Mac Aogáin M and O'Toole RF*. A Step-by-Step Beginner's Protocol for Whole-genome Sequencing of Human Bacterial Pathogens. *Journal of Biological Methods* 2019; 6(1), doi: http://dx.doi.org/10.14440/jbm.2019.276.

Peer reviewed review articles:

‡O'Toole RF* and **Gautam SS**. Limitations of the *Mycobacterium tuberculosis* reference genome H37Rv in the detection of virulence-related loci. *Genomics* 2017; 109(5-6) 471-4, doi: 10.1016/j.ygeno.2017.07.004.

Abstract/conference presentations:

Gautam SS*, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, Globan M and O'Toole RF. Whole-genome Sequencing to Track the Origin and Spread of Tuberculosis in Low Prevalence Setting of Australia. ASM Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines. American Society for Microbiology, 23-26 September 2018, Washington DC area, USA (2018). Available from [Internet]: https://www.asm.org/images/Meetings/NGS_2018/Abstracts/Poster%20B%20Abstracts%20for%20Web.pdf

Gautam SS*, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, Globan M and O'Toole RF. Tracking the Origin and Spread of Tuberculosis in Tasmania

using Whole-genome Sequencing. The University of Sydney, Molecular Microbiology Meeting, 11-12 April 2018, Sydney, Australia (2018).

O'Toole RF*, Mac Aogáin M, Cooley LA, Haug G, Globan M, Fyfe JA, **Gautam SS**. Mapping the origin and spread of *Mycobacterium tuberculosis* complex in the Tasmanian population using whole-genome sequence and epidemiological data analyses. Applied Bioinformatics & Public Health Microbiology, 17-19 May 2017, Hinxton, Cambridge, UK (2017). Available from [Internet]: http://asmicro-2017.p.asnevents.com.au/days/2017-07-05/abstract/43630

Gautam SS*, Mac Aogáin M, Bower JE, Basu I and O'Toole RF. Whole-genome sequence analysis of an outbreak strain of *Mycobacterium tuberculosis*. New Developments in Our Basic Understanding of Tuberculosis (A5). Keystone Symposia in Molecular Biology, 14-18 January 2017, Vancouver, Canada (2017).

Editor reviewed genome reports:

‡**Gautam SS**, Mac Aogáin M and O'Toole RF*. Draft Genome Sequence of the First Confirmed Isolate of Multidrug-Resistant *Mycobacterium tuberculosis* in Tasmania. *Genome Announcements* 2017; 5(44): e01230-17, doi: 10.1128/genomeA.01230-17.

‡Mac Aogáin M, **Gautam SS**, Bower JE, Basu I and O'Toole RF*. Draft Genome Sequence of a New Zealand Rangipo Strain of *Mycobacterium tuberculosis*. *Genome Announcements* 2016; 4(4): e00657-16, doi: 10.1128/genomeA.00657-16.

_____

Sanjay Gautam

Statement of Co-Authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Authors	Affiliations
Sanjay S Gautam	College of Health and Medicine, University of Tasmania, Australia
Micheál Mac Aogáin	Department of Clinical Microbiology, Trinity College Dublin, Ireland
Louise A Cooley	Royal Hobart Hospital, Australia
Greg Haug	Launceston General Hospital, Australia
Kelvin WC Leong	College of Health and Medicine, University of Tasmania, Australia
Rajendra KC	College of Health and Medicine, University of Tasmania, Australia
James E Bower	LabPLUS, Auckland City Hospital, New Zealand
Indira Basu	LabPLUS, Auckland City Hospital, New Zealand
Janet A Fyfe	Mycobacterium Reference Laboratory, Victoria Infectious Diseases Reference Laboratory, Peter Doherty Institute, Australia
Maria Globan	Mycobacterium Reference Laboratory, Victoria Infectious Diseases Reference Laboratory, Peter Doherty Institute, Australia
Ronan F O'Toole	College of Health and Medicine, University of Tasmania, Australia Department of Clinical Microbiology, Trinity College Dublin, Ireland Latrobe University, Australia

Author Details and their Roles

Paper 1. Gautam SS, KC R, Leong KWC, Mac Aogáin M and O'Toole RF. A Step-by-Step Beginner's Protocol for Whole Genome Sequencing of Human Bacterial Pathogens. Journal of Biological Methods 2019; 6(1): e110. DOI:10.14440/jbm.2019.276

Located in chapter 2

Candidate was the primary author and with author 2 and author 3 contributed to the implementation and validation of the research protocol at Respiratory and Clinical Microbiology Group, UTAS. Candidate drafted significant parts of the paper. Candidate contributed approximately 60% to the planning, execution and preparation of the work for the paper.

Author 4 contributed to the conception and design of the protocol.

Author 5 contributed in conceptualization, funding acquisition, investigation, supervision and reviewed and edited the manuscript for publication.

All the authors verified the final version of manuscript.

Paper 2. Gautam SS, Mac Aogáin M, Bower JE, Basu I and O'Toole RF. Differential carriage of virulence-associated loci in the New Zealand Rangipo outbreak strain of Mycobacterium tuberculosis. Infectious Diseases 2017; 49(9) 680-8. DOI: 10.1080/23744235.2017.1330553

Located in chapter 3

Candidate was the primary author and performed next-generation sequencing experiment and preliminary bioinformatics analyses. Candidate and Author 5 conceived and designed the project. Candidate drafted significant parts of the paper. Candidate contributed

approximately 60% to the planning, execution and preparation of the work for the paper.

Author 2 contributed in analysing and validating the bioinformatics analyses.

Author 3 and Author 4 contributed by providing genomic DNA sample and related data.

Author 5 contributed in conceptualization, funding acquisition, investigation, supervision and reviewed and edited the manuscript for publication.

All the authors verified the final version of manuscript.

Paper 3. Mac Aogáin M, Gautam SS, Bower JE, Basu I and O'Toole RF. Draft Genome Sequence of a New Zealand Rangipo Strain of Mycobacterium tuberculosis. Genome Announcement 2016; 4(4): e00657-16. DOI:10.1128/genomeA.00657-16.

Located in chapter 3

Candidate was the co-author and performed next-generation sequencing experiment and preliminary bioinformatics analyses. Candidate and Author 5 conceived and designed the project. Candidate contributed approximately 40% to the planning, execution and preparation of the work for the paper.

Author 1 contributed in submitting the genome sequence to public databases, analyse and validate the bioinformatics analyses.

Author 3 and Author 4 contributed by providing genomic DNA sample and related data.

Author 5 contributed in conceptualization, funding acquisition, investigation, supervision and reviewed and edited the manuscript for publication.

All the authors verified the final version of manuscript.

Paper 4. O'Toole RF and Gautam SS. Limitations of the Mycobacterium tuberculosis reference genome H37Rv in the detection of virulence-related loci. Genomics 2017; 109(5-6) 471-4. DOI: 10.1016/j.ygeno.2017.07.004.

Located in chapter 3

Candidate was the co-author and contributed in drafting of manuscript. Candidate contributed approximately 40% to the planning, execution and preparation of the work for the paper.

Author 1 contributed in conceptualization, funding acquisition, manuscript writing, reviewing and editing.

All the authors verified the final version of manuscript.

Paper 5. Gautam SS, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, Globan M and O'Toole RF. Molecular Epidemiology of Tuberculosis in Tasmania and Genomic Characterisation of its First Known Multi-Drug Resistant Case. PLoS ONE 2018; 13(2):e0192351. DOI: 10.1371/journal.pone.0192351

Located in chapter 4

Candidate was the primary author and performed next-generation sequencing experiment, preliminary bioinformatics analyses and drafted, prepared, reviewed and edited the draft manuscript. Candidate and Author 7 conceived and designed the project. Candidate contributed approximately 65% to the planning, execution and preparation of the work for the paper.

Author 2 contributed in analysis and validation of bioinformatics analyses.

Author 3 contributed in acquisition of fund, genomic DNA sample and patient information.

Author 4 contributed in investigation and data curation.

Author 5 and Author 6 contributed in acquisition of genomic DNA sample and patient information.

Author 7 contributed in conceptualization, funding acquisition, investigation, supervision and reviewed and edited the manuscript for publication.

All the authors verified the final version of manuscript.

Paper 6. Gautam SS, Mac Aogáin M and O'Toole RF. Draft Genome Sequence of the First Confirmed Isolate of Multidrug-Resistant Mycobacterium tuberculosis in Tasmania. Genome Announcement 2017; 5: e01230-17. DOI: 10.1128/genomeA.01230-17.

Located in chapter 4

Candidate was the primary author and performed next generation sequencing experiment and preliminary bioinformatics analyses. Candidate and Author 3 conceived and designed the project. Candidate contributed approximately 60% to the planning, execution and preparation of the work for the paper.

Author 2 contributed in analysis and validation of bioinformatics analyses.

Author 3 contributed in contributed in conceptualization, funding acquisition, investigation, supervision and reviewed and edited the manuscript for publication.

All the authors verified the final version of manuscript.

Paper 7. Gautam SS, Haug G, Cooley LA, Mac Aogáin M and O'Toole RF. Potential cross-border translocation of multi-drug resistant tuberculosis in latent form to Australia. Medical Journal of Australia 2019; 210 (5): 236. DOI: 10.5694/mja2.50044.

Located in chapter 5

Candidate was the primary author and travelled to collect the sample, performed next-generation sequencing experiment, preliminary bioinformatics analyses and drafted the manuscript. Candidate and Author 5 conceived and designed the project. Candidate contributed

approximately 70% to the planning, execution and preparation of the work for the paper.

Author 2 contributed in data curation.

Author 3 contributed in funding acquisition and data curation.

Author 4 contributed in analysis and validation of bioinformatics analyses.

Author 5 contributed in conceptualization, funding acquisition, investigation, supervision and reviewed and edited the manuscript for publication.

All the authors verified the final version of manuscript.

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above ***submitted*** peer-reviewed manuscripts contributing to this thesis:

Signed:

Dr. Ronan O'Toole	Prof. Ben Canny
Primary Supervisor	Head of School
School of Medicine	School of Medicine
University of Tasmania	University of Tasmania
Date: 21 March 2019	22/03/2019

Dedication

This work is dedicated to all the people working in the field of tuberculosis control and management.

And, Professor Dr. Upendra Devkota (1953-2018).

Sanjay Gautam

TABLE OF CONTENTS

Acknowledgements

I owe my deepest gratitude to my primary supervisor, Dr. Ronan O'Toole whose encouragement, guidance and support right from the first day enabled me to envisage the science of tuberculosis and come up with the present thesis. Thank you, sir!!

I am obliged to the participants of my study and thankful to the generous support from Dr. Louise Cooley and Belinda McEwan at Royal Hobart Hospital, Dr. Greg Haug at Launceston General Hospital, Maria Globan and Janet A. Fyfe at VIDRL, Melbourne, James E. Bower and Indira Basu at LabPlus, New Zealand who helped to access samples for the current study. The continuous support of Mr. James Marthick in terms of operating the sequencer and Dr. Micheál Mac Aogáin to perform the bioinformatic analyses allowed me to gain insight into various aspects of next generation sequencing. I am pleased for the co-operation extended by Dr. Phu Truong at Cho Ray Hospital, Vietnam.

This thesis would not have been possible without the support of funding agencies, University of Tasmania and members of the Respiratory and Clinical Microbiology Group at UTAS.

I am indebted to my wife, Ms. Sangeeta Khadka and my parents for their overwhelming support and encouragement. I am obliged to all the well-wishers whose inspiration has always encouraged me to lay the foundation of research and humanity into me.

Sanjay Gautam

Abstract

Tuberculosis (TB) is a major cause of global mortalities causing 1.6 million deaths in 2017 alone due to different species of *Mycobacterium tuberculosis* complex. Although Australia is a low TB burden country, a drop in current annual incidence by approximately 98% is required to achieve the World Health Organization's TB elimination target by 2050. From a public health perspective, challenges exist with regard to tracing the source of TB and identifying factors underlying outbreaks of the disease. The availability of whole-genome sequencing (WGS) has enabled the exploration of the functional and evolutionary genomics of *Mycobacterium tuberculosis*, the detection of drug resistance conferring mutations, and the investigation of epidemiological clusters.

In this thesis, genomic analysis was applied to a dominant outbreak strain of *M. tuberculosis*, the "Rangipo" strain from New Zealand. Whole-genome sequencing of nine isolates representing the Rangipo genotype was performed on an Illumina Miseq platform. The sequence data of each isolate was mapped to the reference *M. tuberculosis* strain H37Rv to generate a consensus genome. There were 700 single locus variants present across all of the outbreak isolates when compared to H37Rv. These included polymorphisms in 12 loci involved in the virulence of *M. tuberculosis*. For example, non-synonymous polymorphism was found in the two-component response regulator *phoR* gene that is essential for the growth of *M. tuberculosis* in macrophage and mice. Furthermore, *de novo* assembly was performed on the sequence reads that did not map to the reference genome

H37Rv. This detected the presence of five additional virulence related genes in the outbreak strain that were absent in H37Rv. These included transcriptional regulator *EmbR2*, molybdopterin cofactor (MoCo) biosynthesis proteins A and B. MoCo is the cofactor for the *narGHI* encoded nitrate reductase which is involved in the adaptation of *M. tuberculosis* in hypoxic conditions and persistence in the guinea pig lungs. These results highlight the presence of additional virulence related genes in the Rangipo outbreak strain that are not present in the reference genome, H37Rv.

Having successfully applied genomic analysis to a New Zealand outbreak strain, attention was then turned to TB in a local setting i.e. Tasmania, where published information on the molecular epidemiology of the disease was limited. I performed the whole-genome sequencing of *M. tuberculosis* isolated in Tasmania and analysed the genomic data together with public health surveillance records. A high proportion (>80%) of TB cases in Tasmania from 2014 to 2016 occurred in overseas born individuals. The whole-genome sequencing data determined the predominance of the East-African Indian lineage 3 of *M. tuberculosis* followed by the Euro-American lineage 4, Indo-Oceanic lineage 1 and East-Asian lineage 2 among TB cases in Tasmania.

Among the lineage 3 isolates, a possible cluster of TB was identified based on the single nucleotide polymorphism (SNP) difference being ≤5 for four of the isolates. Further investigation of the epidemiological data identified that the possible cluster of TB in Tasmania consisted of pulmonary TB cases reported in 2015 in patients originating from Nepal. *In silico* spoligotypes

were generated for the clustered isolates and exactly matched the spoligotype of the dominant lineage 3 genotype, CAS1_Delhi in Nepal. This indicates that the probable origin of the strain of lineage 3 TB cluster cases in Tasmania was the Nepal region.

One of the lineage 2 isolates was collected from a case of extrapulmonary TB which occurred in a 37-year-old male individual originally from Vietnam. The patient had earlier tested positive in interferon gamma release assay in 2016 but did not exhibit the clinical signs and symptoms of pulmonary TB. Following an episode of colitis later in the year, the colon tissue biopsy specimen detected the presence of *M. tuberculosis* in the culture. The isolate was found to be resistant to isoniazid, rifampicin, pyrazinamide and ethambutol and therefore represented the first confirmed case of multi drug resistant (MDR) in Tasmania (TASMDR1). Epidemiological data revealed that a household contact of the Tasmanian MDR-TB patient was diagnosed with pulmonary MDR-TB in Vietnam in 2012. Both the 2016 Tasmanian and 2012 Vietnamese isolates were acquired and upon genome sequencing were found to possess identical high confidence mutations to isoniazid, rifampicin, pyrazinamide, ethambutol and also streptomycin. In addition, the two isolates differed by less than 5 SNPs which is strongly indicative that the two patients were part of the same transmission network. It is highly likely that the Tasmanian case contracted the MDR-TB strain from his household contact in Vietnam and the infection remained in the latent stage before reactivating in the extrapulmonary form in Tasmania in 2016.

In conclusion, this study highlights that differences in the genome content of TB outbreak strains may be undetectable when *M. tuberculosis* sequence data is mapped with a single reference strain. Furthermore, I conclude that the epidemiology of TB in the low prevalence setting of Tasmania has features that resemble TB in other jurisdictions, for example, the presence of the clustered cases and drug resistance.

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations and Symbols

BCG	Bacillus Calmette–Guérin
CD	Cluster of Differentiation
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
COG	Clusters of Orthologous Groups
CXR	Chest X-Ray
DNA	Deoxyribonucleic Acid
EAI	East African/Indian
EPTB	Extra-Pulmonary Tuberculosis
HIV	Human Immuno-deficiency Virus
IGRA	Interferon Gamma Release Assay
IS	Insertion Sequence
LTB	Latent Tuberculosis
MDR-TB	Multi Drug Resistant Tuberculosis
MGIT	Mycobacteria Growth Indicator Tube
MIRU	Mycobacterial Interspersed Repetitive Unit
MTB	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* Complex
NGS	Next-Generation Sequencing
PCR	Polymerase Chain Reaction
PTB	Pulmonary Tuberculosis
SLV	Single-Locus Variant

SNP	Single-Nucleotide Polymorphism
TB	Tuberculosis
TST	Tuberculin Skin Test
VNTR	Variable Number Tandem Repeat
WGS	Whole-genome Sequencing
WHO	World Health Organization
XDR-TB	Extensively-Drug Resistant Tuberculosis

Chapter 1

Tuberculosis: An Introduction

1.1 Introduction

Tuberculosis (TB) is an infectious disease caused by a group of bacteria categorised as the *Mycobacterium tuberculosis* complex (MTBC) that originated from a common African ancestral strain 35,000-15,000 years ago (Brosch et al., 2002; Gagneux, 2018). Based on genomic polymorphism data seven distinct phylogenetic lineages have been identified for the species *M. tuberculosis* (Firdessa et al., 2013; Gagneux et al., 2006). These different lineages of MTBCs differ with respect to their global distribution and in some cases the infectivity, transmissibility and drug resistance (Click et al., 2012; European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis, 2006; Parwati et al., 2010; Reed et al., 2009). The human infection due to MTBC primarily affects the lungs (85% of all new and relapse cases worldwide) and is transmitted via airborne droplets (World Health Organization, 2017). However, a non-transmissible form affecting extrapulmonary body sites, for example, bone, joint, urogenital, ocular, gastrointestinal and other organs (Kulchavenya, 2014) accounted for 15% of global TB incidents in 2016 (World Health Organization, 2017). Patients with pulmonary TB (TB hereinafter) are classified as having active TB, latent TB and subclinical TB (Pai et al., 2016). The first group experience symptoms of TB (for example fever, weight loss and haemoptysis), transmit

infection and are usually detected by sputum smear microscopy and culture-based diagnostic tests. Secondly, patients with latent TB, although infected, do not exhibit symptoms or transmit the infection and are undetectable in culture-based tests (Barry et al., 2009; Desikan, 2013; Esmail et.al, 2014). Thirdly, some patients with sub-clinical TB test positive for tubercle bacilli in their clinical samples and transmit the bacilli but do not demonstrate signs and symptoms associated with tuberculosis (Esmail et al., 2014). With approximately 1 billion deaths in the past 200 years (Ryan, 1994), TB has a significant impact on the public health and global economy. Tuberculosis causes a global loss of approximately 21 billion US dollars per year (Lewis et al., 2018) primarily impacting the gross domestic productivity of high burden countries (Kirigia & Muthuri, 2016).

1.2 Tuberculosis: Epidemiology

TB transmits from person to person via air droplet nuclei particles that contain *M. tuberculosis* complex bacteria produced during coughing, sneezing or talking. They may also be produced during medical procedures, for example, induction of sputum and bronchoscopy (Diagnostic Standards and Classification of Tuberculosis in Adults and Children, 2000). Approximately 3,000 droplet nuclei are produced during a cough or a 5-minute talk (Atkinson et al., 2009) which on average contains 1 -10 bacilli within its size of 1 -5 μM diameter (Lee, 2016). However, exposure does not always result in TB infection and disease. Tuberculosis is less infectious (average transmission of infection from source, 3 -10 people per year) (van

Leth et al., 2008) than measles that infects 90% of its non-immune contacts (Centers for Disease Control and Prevention, 2018). The factors that determine an effective transmission of TB include:- the load of bacilli generated from an infecting source, the concentration of organisms retained in the air after the expulsion, duration of exposure, and the immune status of the host (Diagnostic Standards and Classification of Tuberculosis in Adults and Children, 2000).

Tuberculosis has now emerged as the most common cause of human mortalities due to a single infectious disease accounting for 1.6 million deaths in 2017 (World Health Organization, 2018d) as compared to 0.9 million global deaths due to human immunodeficiency virus (HIV) in the same year (World Health Organization, 2018c). Approximately, 10 million people, including 9 million adults, globally contracted TB in 2017 (World Health Organization, 2018d). The recent global scenario showcases an uneven distribution of TB with countries on the lower side of the economy being affected the most. For example, India, Indonesia and Nigeria together consisted of 39% of total TB incidences worldwide in 2017. The World Health Organization (WHO) South-East Asia region dominated the total share of global TB incidence in 2017 (44%) followed by Africa (25%), the Western Pacific (18%) and minor proportions (13.2%) in Eastern Mediterranean, European and Americas collectively (Figure 1.1) (World Health Organization, 2018d). The annual incidence of TB cases (per 100,000 population) varied from less than 10 in high income countries (for example, Australia) to between 150 -400 in the majority of heavily affected

countries (for example, India) and more than 500 in Mozambique, the Philippines and South Africa (World Health Organization, 2018d).

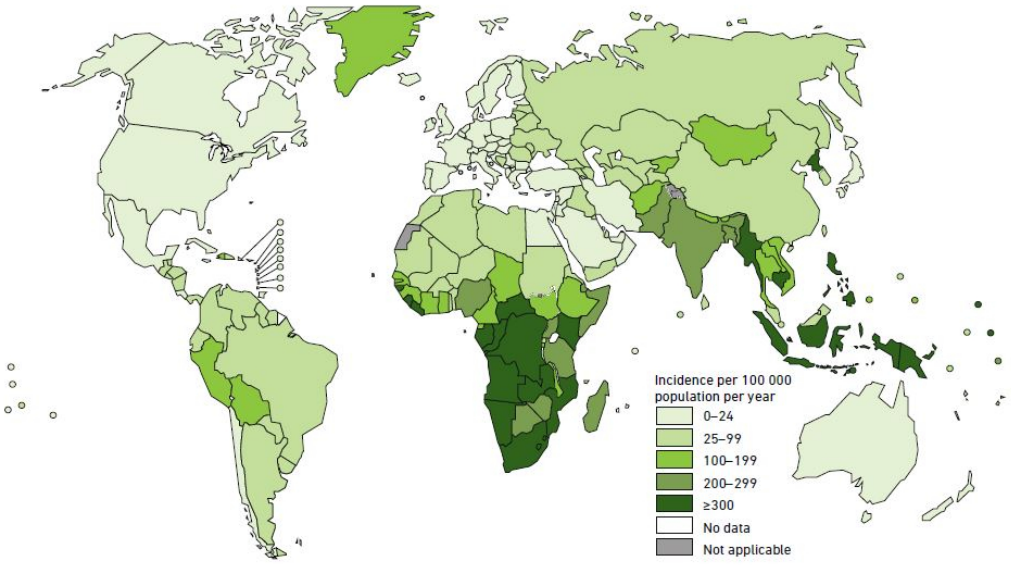


Figure 1.1. Global distribution of Tuberculosis, 2017.

Adapted with permission from the World Health Organization (2018). Copyright World Health Organization.

In terms of treating new cases, a global success rate of 82% was achieved in 2016. In the past 17 years (2000 -2017), 42% deaths due to TB were avoided. However, the emergence of drug resistant TB has added to the challenge in TB control. Approximately, 1.6 million cases in 2017 were reportedly resistant to the first two anti-tuberculosis drugs, isoniazid and rifampicin or mono resistant to rifampicin. India, China and the Russian Federation accounted for almost half of these drug-resistant cases (World Health Organization, 2018d). Although TB has been a continuous threat to mankind, an improvement in terms of reducing global mortality (among HIV

negative patients) by 3% per year is being achieved. The mortality rate in the past 5 years declined fastest (11%) in the WHO European Region in 2016, (World Health Organization, 2018d). The preventive Bacillus Calmette –Guérin (BCG) vaccines average coverage was more than 86% across all WHO regions (World Health Organization, 2017b). The WHO's End TB Strategy and the United Nations' Sustainable Development Goals focus to work "to end the global TB epidemic" for the period of 2016 -2035. This targets for the 80% decrease in TB incidence (new case per year) and 90% reduction in TB related deaths by 2030 in comparison to figures in 2015 (World Health Organization, 2016). To achieve the goal, increased funding in terms of the development of new diagnostics and achieving a technological breakthrough in TB control by 2025 has been advocated (World Health Organization, 2017).

The transmission of TB from an active case is the major source of spread in endemic/epidemic settings and is evidenced by a large number (87%) of new TB cases in the top 30 high burden settings (Geng et al., 2002). In contrast, reactivation of latent TB accounts for most of the active cases (52%) in low TB burden countries and is common in foreign-born residents (Geng et al., 2002). Latent TB (LTB) is a hallmark of insufficient clearance of *M. tuberculosis* from the granulomatous lesion, thereby favouring bacterial persistence and their possible reactivation during an immune suppressed state of the host (Ulrichs & Kaufmann, 2006). A mathematical model estimated the global prevalence of LTB infection to be 1.7 billion people (approximate 1/3rd of the global population) in 2014 and is projected to add 16.5 cases per 100,000 per year in 2035 (Houben & Dodd, 2016).

Therefore, the TB control programs should cumulatively address the burden of active and latent TB to achieve WHO's goal of eradicating the global burden of TB.

1.2.1 Tuberculosis in Australia

The collection of Australian TB notification data started in 1917 (Communicable Diseases Intelligence, 2003). A steady decline in TB notifications was observed between 1918 to 1991 (90.2 versus 3.4 per 100,000 population respectively), with an accelerated decrease starting in 1952 when notification rates reached 55.4 per 100,000 population (Communicable Diseases Intelligence, 2003). Today, Australia is one of the low TB burden countries where 1442 cases (new and relapse) were notified among 24 million people in 2017 (World Health Organization, 2018). Eighty seven percent (n=1255) of TB cases in Australia were detected among those with known HIV status where 27 cases were associated with HIV positive individuals. Sixty two percent had pulmonary TB of which 87% were bacteriologically confirmed (World Health Organization, 2018). There were 25 pulmonary TB cases identified as MDR/RR-TB during 2017 (World Health Organization, 2018). Although the national incidence rate is low in Australia, specific groups of people are at a high risk of developing the disease. For example, the incidence of TB in indigenous Australian is 6 times higher than Australian-born non-indigenous people (Toms et al., 2017). Similarly, overseas born population contribute a significant proportion of TB cases in Australia (86%, n=1151 out of 1339) and international students are more likely to develop TB while they are in Australia (Toms et al., 2017). The most common TB patient's country of

origin (India, Vietnam, Philippines and China) in Australia (Toms et al., 2017) are categorised under WHO's high TB burden nations (World Health Organization, 2015b).

1.3 Tuberculosis: Prevention through Vaccination

For almost 80 years, BCG is the only globally accepted vaccine against tuberculosis. BCG was derived after attenuation of *Mycobacterium bovis* through 230 passages for 13 years (1908 -1921) (World Health Organization, 2013a). In terms of the evolution, BCG emerged as a result of the deletion of the RD-1 locus that encodes virulence related genes (open reading frames including RV3871 to Rv3879c), for example, 10-kDa cultured filtered protein (CFP-10) and 6-kDa early secreted-antigen (ESTAT-6) (Ganguly et al., 2008). The original BCG strain has undergone several passages and its descendants have been named differently, for example, BCG Tokyo, BCG Pasteur and BCG Russia. Genomic variation among BCG vaccine strains globally have been noticed (Abdallah et al., 2015) but its effect on the efficacy in terms of protection was not evidenced in a systematic review including results from 18 clinical trials (Mangtani et al., 2014). The BCG vaccine confers protection against the progression to TB disease (Roy et al., 2014), extrapulmonary TB (Arbelaez, Nelson, & Munoz, 2000) and milliary TB (Trunz, Fine, & Dye, 2006). The protection was high in children who were strictly tested for their prior exposure to mycobacteria (rate ratio (RR), 0.26; 95% confidence interval (CI), 0.18-0.37), or infants (RR, 0.41; 95% CI, 0.29-0.58) as compared to non-strictly tested children (RR, 0.59; 95% CI, 0.35-1.01) (Mangtani et al., 2014). The protection was weaker for the elderly irrespective of their test for previous

exposure to mycobacterial antigen (RR, 0.88; 95% CI, 0.59-1.31 and RR, 0.81; 95% CI, 0.55-1.22, for tested and non-tested respectively) (Mangtani et al., 2014). For, extra pulmonary TB involving meninges and disseminated TB, protective efficacy was higher in tuberculin tested infants and children (RR, 0.1; 95% CI, 0.01-0.77 versus RR, 0.08; 95% CI, 0.03-0.25 respectively) (Mangtani et al., 2014). The poor efficacy of BCG vaccine has been hypothesized to relate to previous exposure to *Mycobacterium* spp. (Black et al., 2002; Brandt et al., 2002). The BCG vaccine is given at birth and offers effective protection for the first 10 years of life (Rodrigues, Mangtani, & Abubakar, 2011). Vaccination is not recommended for protection against primary *M. tuberculosis* infection or revival of latent lung infection (World Health Organization, 2018a) and against leprosy and buruli ulcer (World Health Organization, 2017b). However, BCG is the only available vaccine for TB prevention. World Health Organization recommends an inoculation of a single dose of BCG vaccine at birth to all the infants born in high endemic areas except for those infected with HIV. In low incidence settings (for example, Western Europe, North America and Australia) vaccination should be scheduled to a high-risk group, including neonates immigrating from high risk countries or born to parents with a history of mycobacterial infection (World Health Organization, 2017b).

1.4 Tuberculosis: Risk Factors

The chances of progressing from infection to active TB is highest among infants aged less than a year (pulmonary disease, 30 -40%) (Marais et al.,

2006). The risk lowers in children at their age of 2 -10 years (pulmonary disease, <5%) before rising high (pulmonary disease in 10-20%) after the age of 10 years (Marais et al., 2006). Childhood TB accounted for an approximate 6.9% of notified cases worldwide in 2016 (Swaminathan & Rekha, 2010). The gender wise distribution showed male predominance accounting 64% of total global TB incidences in 2017 (World Health Organization, 2018d). Infection with HIV is one of the most common risk factors for TB where 9% of people living with HIV were detected to have TB in 2017 and the majority (72%) of these cases were concentrated in Africa (World Health Organization, 2018d). In the same year, the death toll due to TB in the HIV-infected population reached 300,000 (23% of global TB mortalities) (World Health Organization, 2018d). Ford and colleagues in their systemic review involving 313,006 adults and 6,182 children with HIV reported TB as a major reason of hospitalization (18% and 10% respectively) and mortality (25% and 30% respectively) (Ford et al., 2015). Although the coexistence of HIV and TB have been widely reported (Friedland, Churchyard, & Nardell, 2007), only 0.5% of the worlds' population live with HIV (World Health Organization, 2018c). Several other factors that increase the risk of acquiring TB in the general population have been identified (Table 1) (Dheda, Barry, & Maartens, 2016). These include health, socio-behavioural and environmental aspects, for example, diabetes, overcrowding, alcohol consumption and exposure to biomass fuel smoke respectively. It is therefore essential to consider all these factors to define the paradigm of tuberculosis.

Table 1. Risk factors of developing tuberculosis and activation of latent infection.

Risk Factors	Risk of developing active TB (folds)	Risk of conversion of latent TB (presumed) to active TB (folds)
HIV	20-40	50-100
Silicosis	3-4	30
Chronic renal failure requiring dialyses	7-50	10-25
Recipient of glucocorticoid treatment	2	4.9
Diabetes	3	2-3.6
Malnourished or underweight	12	2-3
Biomass fuel exposure	2	Insufficient data
Alcohol misuse	3	1.5
Male sex (after adolescence)	2	Insufficient or discordant data
Age	High, <4 and >20 years	2.2-5.0 (when infected before ≤4 years
Malignancy	4-5	1 6 (head and neck carcinoma)
Immunosuppression related to transplant	15-20	20-74
COPD	2 (inhaling corticosteroid)	Insufficient data
Overcrowding and poverty	Increased risk	Increased risk
Recent TB infection (≤ 2 years)	NA	15.0
Apical fibronodular change in chest radiograph	NA	6-19

1.5 Tuberculosis: Clinical Presentations

Tuberculosis was previously named as phthisis, king's evil, consumption, lupus, white plague and captain of all these men of death (Barberis et al., 2017). The clinical manifestation of pulmonary TB onsets gradually, approximately weeks to months after infection (Diagnostic Standards and Classification of Tuberculosis in Adults and Children, 2000). The active form of TB develops in about 5-10% of exposed individuals. Initially, paratracheal lymphadenopathy may develop due to the spread of bacilli through lymphatic channels. The bacteria may reach the pleura of lungs and cause effusion and exhibit symptoms including fever, pleuritic pain and dyspnoea (shortness and difficulty in breathing). The physician may notice poor breath sounds and dullness to percussion as a sign of pleural effusion (Diagnostic Standards and Classification of Tuberculosis in Adults and Children, 2000; Knechel, 2009). The most evident symptom is a non-remitting persistent cough that is reported in 95% cases (Davies, Gordon, & Davies, 2014). However, patients also experience fever (75%), night sweats (45%) and loss in body weight (55%) (Davies, Gordon, & Davies, 2014), progressive fatigue and malaise (Knechel, 2009). In addition, streaks of blood in sputum may be noticed as a result of ruptured cavitary blood vessels in the lungs (Knechel, 2009).

Individuals with latent TB are non-symptomatic and do not transmit the disease (Pai et al., 2016). The latency in TB may convert to active form if viable bacilli in the necrotic tissues reactivate following a loss in the host's immunity, for example, HIV infection, age (≥65 years), progressive diabetes mellitus, kidney failure, poor nutrition, smoking, drug therapy and organ

transplantation (Frieden et al., 2003; Thrupp et al., 2004). People coinfected with HIV and TB may either present with classic symptoms of TB or are poorly symptomatic (Lee et.al., 2000). As the disease progresses from HIV infection to acquired immune deficiency syndrome (AIDS), the chances of developing cavitary lesions are less likely and TB is poorly diagnosed clinically, both by chest radiography and sputum smear examination (Kwan & Ernst, 2011). Therefore, the TB infectious period of HIV positive individuals is lower than those without HIV (2 months versus 11 months respectively) (Corbett et al., 2004). Extrapulmonary TB (EPTB) in at least one site is detected in 60% patients with AIDS as compared to those without AIDS (p<0.001) (Chaisson et al., 1987). The risk of developing EPTB increases with waning immunity as indicated by a drop in CD4 T lymphocyte cell counts, for example, 70% in patients with CD4 cell/µL ≤100 and 28% in those with CD4 cells/ µL count of more than 300 (Jones et al., 1993). This decline in CD4 T lymphocyte count is also associated with the dissemination of TB through the blood circulation. *Mycobacterium tuberculosis* was detected in blood specimen of 49% patients with CD4 cell/µL ≤100 while none in those with the count of more than 300 CD4 cells/ µL (Jones et al., 1993).

In children, poor bacteriological and radiographic evidence complicates the diagnosis of TB (Eamranond & Jaramillo, 2001). Marias and colleagues in their critical review including prechemotherapy literature classified children into high risk group (age <3 years and/or immune compromised) and low risk (age >3 years and/or immune competent). In the former, patients presented with non-remitting and persistent symptoms and occasionally

had an acute onset and frequently progressed to the disease. In the low risk group, conversion from infection to disease was rare but patients clinically presented with symptoms related to the high-risk group (non-remitting and persistent symptoms) (Marais et al., 2006).

1.6 Tuberculosis: Diagnosis

1.6.1 Chest Radiography

The radiological imaging of chest using X-ray (CXR) has long been used to diagnose pulmonary TB (World Health Organization, 2012). The diagnosis of bacteriologically unconfirmed TB is collectively done by assessing the clinical features and CXR report (World Health Organization, 2014a). Tewolde et al. reported the common CXR observations of 159 smear negative pulmonary TB cases where consolidation was detected in a majority of CXR films (40.3%) followed by cavitation (23.9%) and nodular lesions (17.0%) (Tewolde et al., 2015). The lung radiological observation is independent of the duration of infection and shows upper lobe cavitation in adults and immunocompetent host whereas lower lung inflammation, adenopathy and effusion in patients with compromised immune status (Rozenshtein et al., 2015). The utility of CXR in TB screening has been reported from a prevalence study where 5% of 1.2 million immigrants and refugees displayed evidence of TB (World Health Organization, 2012). This shows the applicability of CXR in finding a significant number of TB cases in a pool of healthy population (World Health Organization, 2012). World

Health Organization's guideline to manage latent TB in resourceful countries with less than 100 cases per 100,000 population recommends examining symptoms of TB and perform CXR prior to the initiation of drug therapy (World Health Organization, 2015a). However, in a hospital setting of Nepal, Kumar and colleagues involved 25 physicians to assess 75 CXR films and found a poor specificity (51%) but higher sensitivity (71%) of CXR in diagnosing pulmonary TB (Kumar et al., 2005). The authors highlighted the possibility of over-diagnosis and treatment owing to the poor specificity of CXR in a resource limited setting (Kumar et al., 2005). A recent advent in computed tomography is being considered to be a more sensitive imaging technique to detect TB, lymphadenopathy, bronchogenic spread, differentiating active versus inactive disease and LTB infection more specifically in a patient with higher risk of activation and on drug therapy (Alkabab et al., 2018; Lyu et al., 2011; Piccazzo et al, 2014; Skoura, Zumla, & Bomanji, 2015).

1.6.2 Laboratory Diagnosis

1.6.2.1 Testing for Latent TB Infection

World Health Organization recommends the use of tuberculin skin test (TST) or interferon-gamma release assay (IGRA) depending upon their availability and affordability for the detection of latent TB (World Health Organization, 2015a). These tests are not targeted to diagnose the case or suspected case of active TB. For HIV-TB coinfection, these tests are highly

recommended to detect cases of LTB infection. However, their unavailability should not limit the initiation of therapy in HIV positive individuals or children aged < 5 years with household TB contact (Organization, 2018c). Patients detected positive in either TST or IGRA should undergo CXR before selecting treatment for latent TB, TB or any other lung diseases (World Health Organization, 2015a).

1.6.2.2 Systematic Screening of Active TB

World Health Organization advocates the systematic screening for the early detection of TB and initiation of treatment (World Health Organization, 2013). This includes rapid detection of expected TB cases in specific groups. For example, contacts of active cases, people living with HIV and workers with a history (current or previous) of silica exposure. The guideline also attracts prison inmates, people with fibrosis in the chest (as detected in CXR), residents in high-risk settings or seeking health service living in an area where TB prevalence is ≥100 per 100,000 population, vulnerable people with poor access to health care or in places where the TB prevalence rate is ≥ 1% (World Health Organization, 2013).

The systematic screening involves an initial assessment based on an interview regarding TB symptoms and HIV status. This is then followed by CXR in suspected cases. For adults without HIV, sputum smear examination is followed by a highly accurate molecular test (for example, Xpert MTB/RIF) for both the smear-positive and smear-negative TB cases. Interpretation of these tests results may require re-testing, growing the

bacilli in a culture medium, drug-susceptibility testing and clinical correlation (World Health Organization, 2013). Systematic screening of TB does not include the gold standard culture-based tests due to their longer turnaround time and resource allocation. However, for discrepant screening results and upon the availability of adequate facilities, culture-based identification and drug susceptibility testing is recommended (World Health Organization, 2013).

1.6.2.3 Whole-Genome Sequencing in Diagnosis and Predicting Drug Susceptibility of *M. tuberculosis*

Whole-genome sequencing (WGS) allows the investigation of the entire genomic region of a microorganism (Schwarze et al., 2018). The high throughput next-generation sequencing (NGS) method is based on sequencing by synthesis technology (Buermans & den Dunnen, 2014) and has been applied to generate whole-genome sequence data for the routine characterization (Votintseva et al., 2017), drug susceptibility testing (Witney et al., 2015), genotyping (Amlerova, Bitar, & Hrabak, 2018; Guthrie et al., 2018) and epidemiological investigations of tuberculosis (Gautam et al., 2018). Whole-genome sequencing poses an advantage to ideally identify all the clinically relevant mutations and predict a wide range of drug resistance determining mutations (Wells et al., 2013) as compared to Xpert MTB/RIF and line probe assays which are limited to detect mutations in specific gene locus (Sanchez-Padilla et al., 2015). Furthermore, WGS can identify mutations that confer resistance to new drugs, for example, bedaquiline and

delamanid (Bloemberg, Gagneux, & Böttger, 2015). In terms of clinical application, a multi-centre study involving eight laboratories across Europe and North America reported an accuracy of 93% for species identification and drug susceptibility testing using WGS (Pankhurst et al., 2016). The study linked 15 out of 91 cases to an outbreak and identified a new cluster of TB due to multi-drug resistant strains of *M. tuberculosis*. Furthermore, WGS generated reports 21 days earlier than the existing algorithm, resulting in an annual financial saving of 7% (Pankhurst et al., 2016). Votintseva and colleagues recently reported the utility of WGS in detecting and drug susceptibility testing of *M. tuberculosis* directly from clinical specimen (Votintseva et al., 2017). *Mycobacterium tuberculosis* was identified in all the DNA extracts from 39 out of 40 smear positive respiratory samples (95% to species level) (Votintseva et al., 2017). The genotypic drug susceptibility result was available for 62% (n=24) cases and were all comparable to phenotypic observation. In terms of cost and turnaround time, Illumina MiSeq platform was cheaper than Illumina MiniSeq (£96 versus £198 per sample) but required an additional 28 hours to generate (44 versus 16 hours) the results (Votintseva et al., 2017). Similarly, Brown et al. introduced biotinylated RNA baits to specifically capture *M. tuberculosis* DNA during its extraction from sputum and performed WGS where sequence data of high quality (read depth >20 times and coverage >90%) was obtained from 20 out of 24 smear positive specimens and one smear positive but culture negative case (Brown et al., 2015). The WGS based genotypic drug susceptibility testing was similar to those observed using phenotypic methods (Brown et al., 2015). However, one report suggests potential

contamination due to human genome that leads to the poor prediction of drug resistance when clinical specimen is directly sequenced for *M. tuberculosis* (Doughty et al., 2014). The availability of user-friendly open source tools and databases (for example, PhyResSE (Feuerriegel et al., 2015), TBProfiler (Coll et al., 2015), TGS-TB (Sekizuka et al., 2015) and Mykrobe Predictor (Bradley et al., 2015) to analyses the genomic data of *M. tuberculosis* aids in translating WGS from research to the diagnostic laboratories. However potential limitations in its application to routine diagnostics include cost, technical requirement and standardization of bioinformatic pipelines (Lee & Pai, 2017).

World Health Organization is soon likely to publish a technical guideline addressing the application and interpretation of DNA sequencing technology in TB diagnostics. In addition, availability of a robust description of high confidence mutations of clinical utility that are related to drug resistance in *M. tuberculosis* will likely lead to the extension of WGS to clinical settings (World Health Organization, 2018d).

1.7 Tuberculosis: Pathogenesis

1.7.1 Taxonomic Hierarchy and Microbiology of *M. tuberculosis*

Kingdom	Bacteria
Phylum	Actinobacteria
Suborder	Corynebacterineae
Family	Mycobacteriaceae

Genus	*Mycobacterium*
Species	*Mycobacterium tuberculosis*

The genus, *Mycobacterium* encompasses 170 members reported until 2017 (Forbes, 2017). The genome, seven members of *M. tuberculosis* complex including *M. tuberculosis, M. canettii, M. africanum, M. bovis, M. microti, M. pinnipedii*, and *M. caprae* share 99.9% sequence similarity at 16S rRNA nucleotide level (Brosch et al., 2002; Ernst, Trevejo-Nuñez, & Banaiee, 2007). Based on the DNA-DNA hybridization and average nucleotide identity similarity, Riojas and colleagues have formally proposed reclassification of *M. tuberculosis, M. africanum, M.bovis, M.caprae, M. microti and M. pinnipedi* as the heterotypic synonym of *M. tuberculosis* (Riojas et al., 2017). *Mycobacterium tuberculosis* is not identified in Grams' stain preparation due to the abundant lipid content in their cell envelope (40% cellular dry mass) (Koch & Mizrahi, 2018). The cell wall of *M. tuberculosis* contains a thick layer of peptidoglycan, lipids, glycolipids and polysaccharide, for example, mycolic and mycocerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan (Cole et al., 1998). Furthermore, the tubercle bacilli are acid-fast (i.e. cell wall resists acid decolorization), do not possess flagella for motility, non-spore-forming, require oxygen for survival (i.e. obligate aerobe) and often form beads in the culture media (Dunn, Starke, & Revell, 2016). The bacterium resembles *Mycobacterium leprae* in terms of its intracellular habitat, but unlike leprosy bacillus, *M. tuberculosis* can be retrieved in bacterial culture media (Bhat & Prakash, 2012; Liu et al., 1973). Although *M. tuberculosis* can synthesize all the amino acids and cofactors for enzymatic reactions, it is a slow grower

and requires approximately 22 hours to double its number in a growth medium (Koch & Mizrahi, 2018; Roetzer et al., 2013) as compared to 20 minutes for *Escherichia coli* (Sezonov, Joseleau-Petit, & D'Ari, 2007). This has been related to the slower polymerization rate (20 times less as compared to *Escherichia coli*) of a protein *FtsZ* in *M. tuberculosis* that is responsible for cell division and cell wall biosynthesis (Hett & Rubin, 2008).

1.7.2 Evolution and Molecular Clock of *M. tuberculosis*

The modern *M. tuberculosis* complex strains have lost several repeat regions during the evolutionary process (Arnold et al., 2006). These deletions may be a non-functional phenotypic marker or potentially related to the survival and expansion of the bacillus (Arnold et al., 2006). *Mycobacterium africanum* type 1, the ancestor of *M. microti* and *M. bovis* lost spacer 39 and differed from *M. tuberculosis* that did not possess spacer 34. This *M. tuberculosis* that is absent for specific deletion region (TbD1) emerged as modern strains (for example H37Rv, Beijing) and those that were deleted for spacers 29-32 and 34 diverged as ancient strains (for example atypical *Mycobacterium*) (Arnold et al., 2006). Although the members of *M. tuberculosis* complex are closely related in terms of their 16S rRNA composition (99.9% similarity at the nucleotide level), they significantly differ in ecotypes and infectivity (Brosch et al., 2002). For example, *M. tuberculosis, M. africanum* and *M. canettii* are exclusively human pathogens whereas *M. microti*'s habitat is in rodents and *M. bovis* has a wide range of host tropism (Brosch et al., 2002).

The concept of the *M. tuberculosis* complex molecular clock is based on estimating the number of mutations across the whole-genome per unit time (Barbier & Wirth, 2016). This has been important in studying the evolution of *M. tuberculosis* and genomic variations in *M. tuberculosis* acquired over a period of time (Arnold, 2007). Different studies have reported the occurrence of 0.3-0.5 single nucleotide polymorphism per year/genome of *M. tuberculosis* (Bryant et al., 2013; Roetzer et al., 2013; Walker et., 2013). This corresponds to $1*10^{-7}$ substitutions per nucleotide per year (Barbier & Wirth, 2016) which is lower than *Staphylococcus aureus* (Nübel et al., 2010)*, Escherichia coli* (Wielgoss et al., 2011) and mis -match repair lacking *Helicobacter pylori* (Didelot et al., 2013) ($2*10^{-6}$, $5*10^{-6}$ and $1*10^{-5}$ respectively). Based on the changes in the IS*6110* element, it is estimated that two strains of *M. tuberculosis* separate apart in 3.2-8.7 years (de Boer et al., 1999; Warren et al., 2002; Warren et al., 2002). However, this rate of genetic change differs between the lineage types of *M. tuberculosis* (Kato-Maeda, Metcalfe, & Flores, 2011). The estimation of mutation rate may be affected by the intra-chromosomal variations, lineage difference and stages of infection (for example active versus latent). This calculation may be further complicated by the presence of highly variable repetitive regions prone to mutation (for example, PE-PGRS, PPE and ESX families) and their possible omission during the genomic data analyses (Barbier & Wirth, 2016). The molecular clock estimate of *M. tuberculosis* is being increasingly applied in molecular studies of TB, particularly to establish the epidemiological links between the cases, trace the source and their contacts

to define the events of TB transmission (Bryant et al., 2013; Roetzer et al., 2013; Walker, et al., 2013).

1.7.3 Lineage Classification of *M. tuberculosis*

The availability of whole-genome sequencing has revealed considerable genomic variations among/within the species of *M. tuberculosis* (Fleischmann et al., 2002; Gautam et al., 2017; Ioerger et al., 2010; Zheng et al., 2008). This resulted in large sequence polymorphism (LSPs) and single nucleotide polymorphism (SNPs) as an unambiguous and robust method for phylogenetic classification of *M. tuberculosis* (Filliol et al., 2006; Kato-Maeda et al., 2001). The less likelihood of occurrence of SNP, horizontal gene transfer and re-acquisition of deleted regions in *M. tuberculosis* complex makes SNP and LSP based evolutionary classification more ideal and informative (Filliol et al., 2006; Gagneux et al., 2006; Gutacker et al., 2002; Mostowy et al., 2002).

Mycobacterium tuberculosis has been categorised into different genetic lineages. Gagneux and colleagues' classification into six lineages (L1 - L6 respectively) was based on the presence of large sequence polymorphisms, for example RD239 (L1, Indo-oceanic), RD105 (L2, East Asian), RD750 (L3, East African-Indian), *Pks15/1* △7bp (L4, Euro-American), RD711 (L5, West African I), RD702 (L6, West African II) (Gagneux et al., 2006). Recently, Lineage 7 *M. tuberculosis* "Aethiops vetus" (Nebenzahl-Guimaraes et al., 2016) has been reported among the strains isolated in Ethiopia (Comas et al., 2013; Firdessa et al., 2013). These Ethiopian strains were

phylogenetically closer to L1 East African-Indian (EAI), possessed TbD1 region, RvD1-RvD5 and RD239 (or LSP 239) and had deletions in *mmp*19, *IppH*, *IppO/sseB*, Rv3467/*rmlB2/mhpE* (Nebenzahl-Guimaraes et al., 2016). Different lineages of MTBCs are either geographically restricted (for example L5, L6 and L7) or are widespread (for example L2 and L4) (Figure 1.2) (Gagneux, 2018). I will follow the Gagneux lineage classification hereinafter.

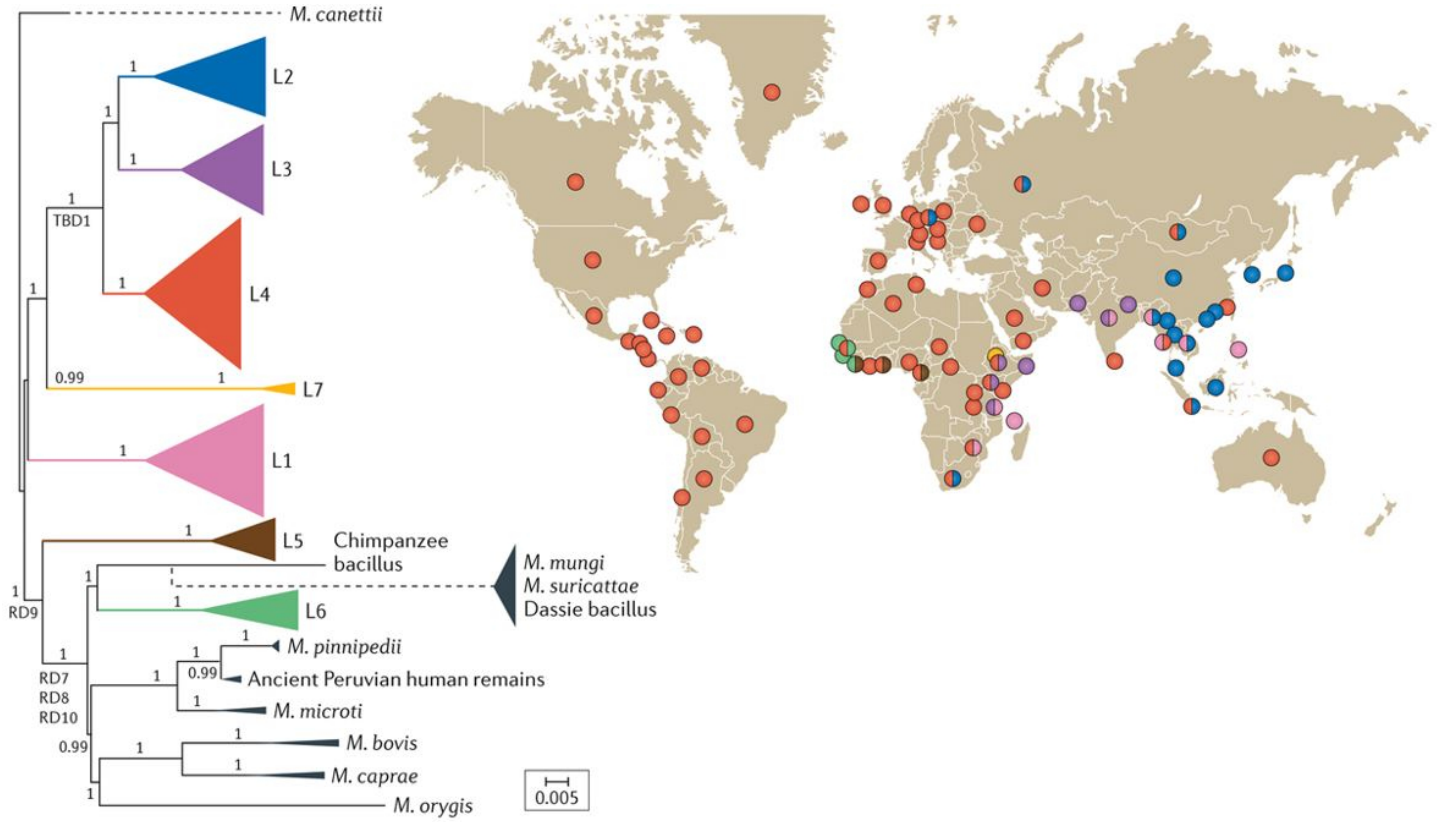


Figure 1.2. Genomic changes during the evolution of MTBC and global distribution of its different lineage types.

The colour codes in the map and phylogenetic tree represent each lineage of *M. tuberculosis* as indicated in the later. The triangles represent collapsed branches (for clarity) of the main *M. tuberculosis* complex lineages.

Adapted with permission from Gagneux, Nat Rev Microbiol. (2018). Copyright Springer Nature Limited.

Various studies have reported the dominant presence of lineage of *M. tuberculosis* circulating in their country of origin among overseas born individuals. Yen and colleagues in their New Zealand based study predominantly detected lineage 2 among patients born in East Asia (34/46, 73.9%) (Yen et al., 2013). Gagneux and colleagues hypothesized that "lineages that are rare in a specific human population are not adapted to transmit and cause secondary cases in this specific human population" (Gagneux et al., 2006). Furthermore, the authors reported that patients born in the US to Chinese and Filipino parents were detected with strains common in China and Philippines, i.e. East-Asian and Indo-oceanic respectively (p <0.001) respectively (Gagneux et al., 2006). This sympatric combination between a specific lineage of *M. tuberculosis* complex and geographic origin of the patient has been associated with clinic-epidemiology of tuberculosis (Gagneux et al., 2006) and may be related to the presence of latent TB acquired in the home country prior to their arrival in the new setting. However, some lineages, for example, lineage 4 is globally distributed but is restrictive in terms of geographic presence of its sub-lineages (Stucki et al., 2016).

1.8 Tuberculosis: Chemotherapy and Drug Resistance

1.8.4.1 Therapeutic Approach

Tuberculosis is a preventable and curable disease (Reichman, 1997). Ironically, after 25 years of its declaration of global emergency, TB is now a

leading cause of deaths (World Health Organization, 2018d). The aim of TB therapy is to completely treat the disease ensuring no reversions, prevent mortalities, disrupt the transmission and prevent the development of drug resistance (du Toit, Pillay, & Danckwerts, 2006). It was during 1944 -1976 the major anti-TB drugs (for example, isoniazid, streptomycin, rifampicin) of current TB treatment regime were introduced (Chaisson, 2003). These antibiotics not only treat the disease but also rapidly decrease the infectiousness of the sources thereby limiting the spread of *M. tuberculosis* (Schwartzman & Menzies, 2000). Isoniazid is lethal in the initial stage of infection killing almost 95% bacilli within the first two days whereas rifampicin is active against persisting mycobacteria in the later stage of infection (Joshi, 2011). The current WHO guideline for the treatment of drug susceptible TB includes 6 months of rifampicin-based therapy which contains an intensive phase of 2 months (Isoniazid (H), Rifampicin (R), Pyrazinamide (Z) and Ethambutol (E)) followed by 4 months of continuation of H and R (World Health Organization, 2017a). Treatment of TB, therefore, involves one of the longest regimes directed against infectious disease (Namasivayam et al., 2017). In terms of HIV infection, antiretrovirals should be administered to all TB patients irrespective of their CD4 T lymphocyte cell count and TB treatment should be done for 6 months. Management of tubercular meningitis should include addition of corticosteroid over the period of 6-8 weeks. Children should be treated with the same regime as for an adult, i.e. 2 months HRZE followed by 4 months HR (World Health Organization, 2014b). In addition to the chemotherapy, provisions for patient care and the directly observed treatment administered by trained personnel

at home or community should be ensured (World Health Organization, 2017a).

To treat latent TB in both the adults and children in high and low burden settings, six months monotherapy with isoniazid is recommended. In high burden areas an alternative three months daily isoniazid and rifampicin (for individuals aged <15 years) or three months weekly isoniazid and rifapentine for adults and children should be offered. In low incidence settings, treatment regime including nine months isoniazid or three to four months isoniazid and rifampicin or three months weekly isoniazid and rifapentine or three to four months of rifampicin monotherapy is recommended. People living with HIV should receive an isoniazid preventive treatment (36 months for adults and adolescents and for 6 months to children ≥12 months, infants <12 months and children with successful TB treatment) (World Health Organization, 2018e). The Centers for Disease Control and Prevention recommends 9 months of isoniazid chemoprophylaxis for immigrants with latent TB (Centers for Disease Control and Prevention, 2012).

In terms of drug development, as of 2018, 20 new anti-TB drug candidates are in different phases of development (World Health Organization, 2018d). This includes two drugs (bedaquiline and delamanid) that have been accelerated with regulatory conditions for TB treatment (World Health Organization, 2013b; World Health Organization, 2014c). In addition, seven drugs have been repurposed for their inclusion in treatment and include

clofazimine, linezolid, levofloxacin, moxifloxacin, nitazoxanide, rifampicin (high dose) and rifapentine (World Health Organization, 2018d).

1.8.4.2 Tuberculosis: Drug Resistance

Drug resistant TB is emerging as a global threat with a record number of 160,684 multi-drug resistant/rifampicin resistant (MDR/RR) cases in 2017 (an approximate 5% increase since 2016) (World Health Organization, 2017; World Health Organization, 2018d). The incidence of MDR/RR-TB was highest in previously treated cases (18%) than in new cases (3.5%) in 2017. India (24%), China (13%) and Russian Federation (10%) collectively accounted for 47% of global MDR/RR-TB cases (World Health Organization, 2018d). In terms of therapy, low success rate was recorded (55% globally) and a big gap was realized in terms of enrolment for treatment with second-line drugs (enrolment proportion, 25% in 2017 versus 87% in 2016) mainly in India and China (40% of the total gap) (World Health Organization, 2018d).

Mycobacterium tuberculosis resistant to at least the two first line anti-TB drugs, isoniazid and rifampicin is defined as multi-drug resistant isolate. These isolates, if they retain resistance to at least one of the second line injectable drugs (kanamycin, amikacin or capreomycin) and a fluoroquinolone, are regarded as extensively-drug resistant *M. tuberculosis* (World Health Organization, 2018d). The drug resistance TB may be acquired when the susceptible isolate develops to a resistant form within the host or due to the spread of the drug resistant strain itself (Joshi, 2011). Drug resistance in *M. tuberculosis* results from chromosomal mutations

(Gygli et al., 2017) and develops as a result of the accumulation of drug resistant progeny (Veen, 1995) rather than the horizontal transfer of resistance determining genes (Cole et al., 1998), plasmids or mobile elements (Gygli et al., 2017). The development of drug resistance in *M. tuberculosis* has been described as a "man-made problem" (Kochi, Vareldzis, & Styblo, 1993) arising due to poor chemotherapy and a result of "careless care" (Raviglione, 2006). Drug resistance in TB can develop during treatment for active TB due to "inadequate clinical care or drug management", for example, incomplete treatment, prescription of the wrong regime in terms of type and dose or unavailability and/or poor quality of drugs (World Health Organization, 2018g). Liang and colleagues identified a higher number of re-treated patients acquiring MDR-TB (5.48 fold higher than new cases) and improper drug therapy as a factor for the development of MDR-TB (Liang et al., 2012). Similarly, age, living in a single roomed house, previous treatment and a history of HIV infection has been identified as independent predictors for MDR-TB (Workicho, Kassahun, & Alemseged, 2017). In addition, patients felt that poor economy, adverse effect of drugs and lack of knowledge as being factors associated with MDR-TB development (Liang et al., 2012).

Mycobacterium tuberculosis is intrinsically resistant to some of the commonly used anti-bacterial drugs, for example, beta lactams (Abraham et al., 1941). However, the acquisition of chromosomal mutations that results in *M. tuberculosis* drug resistance are clinically correlated. These mutations may include single or multiple events that confer resistance of *M. tuberculosis* to anti- tuberculosis drugs (Table 2) (Gygli et al., 2017).

Table 2. Drug resistance determining mutations in *M. tuberculosis.*

Antibiotics	Gene target	Mechanism of resistance	Reference
Rifampicin	*rpoB*	Drug target alteration	(Telenti et al., 1993)
Isoniazid	*katG*	Abrogated drug target	(Heym et al., 1995)
	inhA	Drug target alteration	(Banerjee et al., 1994)
	inhA promoter	Drug target overexpression	(Morlock et al., 2003)
Ethambutol	*embB*	Drug target alteration	(Sreevatsan et al., 1997)
Pyrazinamide	*pncA*	Abrogated prodrug activation	(Konno, Feldmann, & McDermott, 1967)
Ethionamide	*inhA*	Drug target alteration	(Banerjee et al., 1994)
	inhA promoter	Drug target overexpression	(Morlock et al., 2003)
	ethA	Abrogated prodrug activation	(Morlock et al., 2003)
Fluoroquinolones	*gyrA/B*	Drug target alteration	(Takiff et al., 1994)
Streptomycin	*Rrs*	Drug target alteration	(Meier et al., 1994)
	rpsL	Drug target alteration	(Meier et al., 1994)
Amikacin	*Rrs*	Drug target alteration	(Alangaden et al., 1998)
Kanamycin	*Rrs*	Drug target alteration	(Alangaden et al., 1998)
	eis promoter	Overexpression of drug inactivating enzyme	(Kambli et al., 2016)
Capreomycin	*Rrs*	Drug target alteration	(Maus, Plikaytis, & Shinnick, 2005)

	tlyA	Abrogation of drug target methylation	(Maus, Plikaytis, & Shinnick, 2005a)
P-aminosalicylic acid	*thyA*	Drug target bypassing	(Zhao et al., 2014)
	*Fol*c	Abrogation of prodrug activation	(Zhao et al., 2014)
Cycloserine	*Ald*	Overabundance of drug target substrate	(Desjardins et al., 2016)
	Alr	Drug target alteration	(Desjardins et al., 2016)
	alr promoter	Drug target overexpression	(Desjardins et al., 2016)
Bedaquiline	*atpE*	Drug target alteration	(Huitric et al., 2010)
	Promoter/*mmpR*	Overexpression of efflux pump Mmpl5	(Hartkoorn, Uplekar, & Cole, 2014)
Linezolid	*rplC*	Drug target alteration	(Beckert et al.)
	Rrl	Drug target alteration	(Hillemann, Rüsch-Gerdes, & Richter, 2008)
Delamanid/Pretomanid	*Ddn*	Abrogation of prodrug activation	(Haver et al., 2015)
Clofazimine	Promoter/*mmpR*	Overexpression of efflux pump Mmpl5	(Hartkoorn et al., 2014)

The World Health Organization has issued general principles to guide the treatment of MDR-TB (World Health Organization, 2014). This recommends the treatment of MDR-TB with at least four second line drugs including an injectable (for example, amikacin, capreomycin, and kanamycin) and pyrazinamide even if evidence of effectiveness is not clear with any of these drugs. In case the recommended drugs meet a criterion of “likely to be effective”, treatment should include ethionamide or prothionamide. The treatment course may extend up to 20 months for new MDR-TB cases and is modifiable with respect to culture conversion and patients’ reaction to the drugs (World Health Organization, 2014). This longer duration of therapy may add up the additional challenge in TB control, for example, poor compliance and treatment failure leading to continued TB transmission (Namasivayam et al., 2017). The World Health Organization has introduced a shorter MDR-TB regimen including, 4-6 months of kanamycin-moxifloxacin-prothionamide-clofazimine-pyrazinamide-isoniazid$_{\text{high-dose}}$ - ethambutol followed by 5 months of moxifloxacin-clofazimine - pyrazinamide-ethambutol (World Health Organization, 2015). However, this is not recommended for those displaying resistance to shorter regime drugs (excluding isoniazid) or have been exposed to TB for ≥1 month or developed intolerance to these drugs or for the treatment of extra-pulmonary TB and during pregnancy (World Health Organization, 2015). In terms of HIV coinfection, antiretrovirals should be added to the MDR-TB drug regime irrespective of immune status based on the CD4 T lymphocyte cell count (World Health Organization, 2014). After 40 years, WHO issued an interim guideline for inclusion of the new anti-TB drug, bedaquiline in combination

therapy for MDR-TB (World Health Organization, 2013b). Bedaquiline together with delamanid was “justified” to treat TB in those patients left with fewer options of the drug of choice (Andrey et al., 2017).

1.9 Epidemiological Typing of *M. tuberculosis*

Typing of bacterial populations enables one to study their relationship within the members of the same species or wider category. This can further be applied in epidemiological studies, for example, tracing the source of infections/outbreaks, identify transmission chains either in the community or hospital settings, determine pathogenic strains, designing surveillance strategies and evaluating control measures (MacCannell, 2013; Struelens, 1998). Conventionally, bacterial species have been typed based on their phenotypic properties, reactions towards antigen, antibody or biochemicals, and susceptibilities towards phage and antibiotics (Foxman et al., 2005).

The epidemiological typing of *M. tuberculosis* is an important tool for the tuberculosis control strategy (Cannas et al., 2016) particularly due to the global distribution tuberculosis, airborne nature of its spread, clonality in origin, lineages and sub-lineages categories of infecting pathogen, prolonged latency of infection, varying efficacy of BCG vaccination, inapparent clinical signs and potential to cause outbreaks. Here, I will discuss the utility of common molecular epidemiological typing methods, including whole-genome sequencing for the surveillance of tuberculosis.

1.9.1 IS*6110* Restriction Fragment Length Polymorphism

The classification based in the detection of number and position of fragments (generated by restriction enzyme *Pvu*II) of the highly conserved mobile genetic element, IS*6110* that can get inserted anywhere in the *M. tuberculosis* genome forms the basis of the first molecular typing method of *M. tuberculosis*, IS*6110* Restriction Fragment Length Polymorphism (RFLP) (van Embden et al., 1993). The bands of the fragments are compared with the reference fingerprint to standardise the observation across different laboratories (Supply, 2015). This method has been useful in outbreak investigation (van Soolingen et al., 1999), cross-contamination detection in the laboratory (Van Duin et al., 1998) and differentiating exogenous reinfection versus endogenous reactivation (Small et al., 1993). Moreover, IS*6110* RFLP is applicable to the retrospective analyses as it requires the growth of mycobacteria in a growth medium (Supply, 2015). Other limitations include poor standardization due to the complex genome of *M. tuberculosis* (Supply, 2015), bias due to the variable number of inserts (Mazars et al., 2001) or potential absence of IS*6110* copies (Dale, et al.,1999) and technical demands (Cannas et al., 2016).

1.9.2 Spacer Oligotyping (Spoligotyping)

Spoligotyping is a PCR based hybridization technique that detects the presence or absence of spacer sequences within direct repeat locus (a member of clustered, interspaced short palindromic repeat loci within a bacterial family) of *M. tuberculosis* (Kamerbeek et al., 1997). The direct

repeat locus is amplified and hybridised using a reverse line technique to generate 43-bit barcode (indicating presence or absence of spacer) which is compared with the 43 unique spacer references in *M. tuberculosis* (Kamerbeek et al., 1997). The data are generated as a binary code or an octal designation and is classified into clade and phylogenetic group using the SITVIT database (Demay et al., 2012). This is a cost-effective method that can identify different lineages of *M. tuberculosis* (Supply, 2015). However, similarity in the patterns among distantly related sub-strains leads to poor resolution of the method, for example, members of Beijing strains are grouped as the same spoligotype (Kremer et al., 2004). In addition, the phylogenetic inference based in spoligotype pattern may be unclear due to homoplasy (Comas et al., 2009). Therefore, spoligotyping is not recommended to establish epidemiological links between cases, particularly during outbreak investigations (Barnes & Cave, 2003).

1.9.3 Mycobacterial Interspersed Repetitive Unit (MIRU) Variable Number Tandem Repeat (VNTR)

The MIRU-VNTR is based on the amplification technique that uses primers specific to flanking regions to multiply tandem repeat sequences followed by characterization of *M. tuberculosis* depending upon the number of repeats in each region indicated by the size of amplicons (Supply et al., 2006). The information is deduced as a globally standardised code containing 24-number types which can be analysed (for a phylogenetic relationship, genotype and lineage prediction) using the open source

database MIRU-VNTR*plus* (Weniger et al., 2010). This method identifies clusters of cases and ongoing TB transmission based on the similarity of MIRU-VNTR alleles distributed over 24 loci (Allix-Beguec et al., 2014; Savine et al., 2002; Supply et al., 2006). The stability of MIRU-VNTRs allows the monitoring of chronic cases over a period of time (Savine et al., 2002). Furthermore, the detection of two alleles in two or more MIRU-VNTR loci in an isolate indicates potential contamination or mixed infection (Allix, Supply, & Fauville-Dufaux, 2004). In terms of resolution, MIRU-VNTR possesses more discriminatory power than spoligotyping and is less prone to the effect of homoplasy (Cardoso et al., 2011; Wirth et al., 2008). McLernon and colleagues recommend the better performance of MIRU-VNTR and spoligotyping collectively as compared to MIRU-VNTR alone (McLernon et al., 2010). However, MIRU-VNTR does not offer a distinction between closely related members of the *M. tuberculosis* Beijing family (Roetzer et al., 2011). Furthermore, the resolution of VNTR typing is compromised while investigating an established epidemiological link (confirmed and presumed), (de Beer et al., 2013).

A common limitation of the above-mentioned methods is their inability to detect variations in the regions other than the small targeted segments within the whole-genome of *M. tuberculosis*. It is plausible that variations can independently occur in other areas of the genome leading to increased genetic distances between the isolates (Supply, 2015). In addition, these techniques offer poor evidence to the measurement of inter-species and intra-species evolutionary distances between the *M. tuberculosis* isolates (Filliol et al., 2006).

1.9.4 Whole-Genome Sequencing

Whole-genome sequencing allows the examination of the entire region of a genome and facilitates in detecting the variations of possible epidemiological importance (Koser et al., 2012). Single nucleotide polymorphism (SNP) in *M. tuberculosis* genome occurs once in every 3kb of DNA sequence but is the most common form of genetic variation besides insertions/deletions events (Comas et al., 2010). Therefore, SNP based typing can potentially represent a “gold standard” method for molecular typing of *M. tuberculosis* (Filliol et al., 2006). The genome of *M. tuberculosis* is highly conserved which is evidenced by the presence of small number of SNPs between the reference H37Rv and clinical CDC1551 strain (n=1075) (Fleischmann et al., 2002) and is characterized by high clonal population structure (Liu et al., 2006). The availability of molecular clock data based in SNP analyses indicating the emergence of 0.3 to 0.5 SNPs per genome per year has allowed the prediction of epidemiological links between the cases (Bryant et al., 2013; Roetzer et al., 2013; Walker, et al., 2013). Furthermore, a cut-off of ≤5 SNPs between the isolates has been increasingly used to define the TB transmission network (Gautam et al., 2018; Walker et al., 2013). However, this should be interpreted in relation to available epidemiological data as SNP differences may be affected by factors including- sampling time, the incidence of TB and genetic similarities between the strains in a geographical area (Hatherell et al., 2016; Migliori et al., 2008). The lesser likelihood of SNPs to converge adds additional advantage over spoligotyping and MIRU-VNTR in analysing the phylogenetic relationship between the closely related isolates (Schork,

Fallin, & Lanchbury, 2000). Jajou et al. compared WGS and MIRU-VNTR to identify clustered cases and found that WGS is approximately two times more effective than the later in identifying epidemiologically related cases (57% versus 31% respectively) (Jajou et al., 2018).

The WGS based transmission tracking of TB in both high (Bjorn-Mortensen et al., 2016) and low burden countries (Gurjav et al., 2016) have been documented. This has also been used to detect mixed TB infections (Gan et al., 2016; Sobkowiak et al., 2018) and determine microevolution (Kato-Maeda et al., 2013), both at inter and intra-patient level (Pérez-Lago et al., 2014). Whole-genome sequencing identified identical MIRU-VNTR genotypes which along with the analyses of patient's social networks detected the presence of a common progenitor during a TB outbreak that existed before the onset of transmission (Gardy et al., 2011). This outbreak was further related to a social event of the use of crack cocaine (Gardy et al., 2011). Furthermore, WGS allows to analyse the functional genomics of *M. tuberculosis*, for example, Colangeli and colleagues analysed the whole-genome of *M. tuberculosis* to report the growth characteristics and acquisition of drug resistance determining polymorphisms during the latent phase of infection (Colangeli et al., 2014). The application of WGS further extends in predicting *M. tuberculosis* drug susceptibility (Walker et al., 2015), infer phylogenetic relationships (Gautam et al., 2018), evolutionary history (Wollenberg et al., 2017) classify strains into lineages and sub-lineages (Schleusener et al., 2017) and the development of TB diagnostic tools (Manson et al., 2017).

1.10 Concluding Remarks for the Literature Review and Questions for the Present Thesis

Tuberculosis is a major global health crisis, especially in low income countries, but poses a significant threat in specific population groups in developed nations, for example, immigrants and socio-economically disadvantaged communities. Australia has endorsed the WHO's Global strategy and targets for TB prevention, care and control after 2015. To achieve the goal of reducing TB incidence and mortality by 90% and 95% respectively by 2035, a drop by approximately 98% is required in the current annual TB incidence rate in Australia. In terms of diagnostics, the available TB detection and epidemiological typing methods pose limitations in terms of their sensitivity, specificity or methods restricted to specifically analyse a particular gene locus. Whole-genome sequencing is being increasingly developed as a tool to collectively diagnose and perform functional and epidemiological genotyping *M. tuberculosis*.

Thus, the **aims** of my thesis were designed to utilize whole-genome sequencing to explore the genome of *M. tuberculosis* and further implement the technology to study the molecular epidemiology of TB. The aims are summarized as;

1. To apply whole-genome sequencing as a tool for functional genotyping of an outbreak strain of *M. tuberculosis*.
2. To apply whole-genome sequencing as a tool to study the molecular epidemiology of TB in Tasmania.

3. To trace the origin and track the transmission of the first case of multi-drug resistant *M. tuberculosis* detected in Tasmania.

1.11 Hypothesis

The hypothesis for the genomic study of New Zealand TB outbreak "Rangipo" strain relied on the existing report of their high transmissibility. I hypothesized that the genome of Rangipo strain contains additional features that can potentially be related to its increased virulence. In addition to the functional analyses of *M. tuberculosis* "Rangipo" genome, I hypothesized that WGS can be applied for epidemiological typing of tuberculosis in a low TB burden setting. The availability of genome sequencing facility at the University of Tasmania and bacterial samples from institutions in Australia and overseas allowed me to test my thesis hypothesis for the present thesis.

Chapter 2

Method: Whole-Genome Sequencing and Data Analyses

2.1 Introduction

Using Sanger sequencing, the Human Genome Project cost approximately USD $2.7 billion and took 13 years to produce the first human genome sequence (International Human Genome Sequencing Consortium 2004; Gyles, 2008). Today, a human genome can be sequenced in a matter of days for less than USD $1,500 on a single next-generation sequencing (NGS) machine (Wetterstrand, 2018). This change in throughput and the per-base cost has transformed the use of DNA sequencing in biomedical research and is being translated into an expanding number of ways into medicine. NGS is increasingly been applied to understanding and managing infectious diseases. This includes the sequencing of microbial genomes for the purposes of laboratory identification of infectious agents (Lecuit & Eloit, 2014), detection of antibiotic resistance markers (Walker et al., 2015), and the public health surveillance of epidemiological clusters and outbreaks (Harris et al., 2013). Examples include its deployment in public health surveillance and control of community cases of *Escherichia coli* (Grad et al., 2012), *Campylobacter jejuni* (Llarena, Taboada, & Rossi, 2017), *Legionella pneumophila* (Reuter et al., 2013) and *Mycobacterium tuberculosis* (Gautam et al., 2018) disease, or global and regional epidemics caused by influenza (McGinnis, Laplante, Shudt, & George, 2016), Ebola (Park et al., 2015), and Zika (Metsky et al., 2017) viruses. It has also been utilised to

track the source and spread of healthcare-associated infections caused by *Staphylococcus aureus* (Köser et al., 2012), *Pseudomonas aeruginosa* (Davis et al., 2015), *Acinetobacter baumannii* (Halachev et al., 2014), and *Enterococcus faecium* (Leong et al., 2018) to guide the infection prevention and control strategies in health care settings.

In addition to whole-genome sequencing (WGS), whole exome, transcriptome (RNA-Seq), bisulphite methylome, and metagenomic sequencing, NGS can be directed to the detection of specific genes or mutations associated with human disease through targeted-panel amplicon screening. However, barriers remain with regard to establishing NGS in a laboratory for the first time and this hinders its uptake in clinical microbiology and other settings. One of these challenges is the lack of a simplified step-by-step protocol that can be picked up by laboratory personnel with no prior training or experience in NGS and used to generate reliable, high quality sequence data. Illumina dye-sequencing is currently considered the gold standard internationally in terms of the read depth and base-calling accuracy, genome coverage, scalability, and the range of sequencing applications it delivers.

In this work, I produced an easy-to-follow, step-by-step NGS protocol with consistent genome coverage and average read depth that was applicable to a range of bacterial pathogens i.e. Gram-positive vancomycin-resistant *Enterococcus faecium*, Gram-negative non-typeable *Haemophilus influenzae*, and acid-fast high-GC content *Mycobacterium tuberculosis*. This protocol can be used to generate the Illumina-based WGS data for clinical

isolates of bacterial pathogens of importance to human health. This laboratory procedure generated Fastq reads from the sequencer within three days of start. I modified a number of the DNA extraction steps to obtain a sufficient quantity of contamination free template. Similarly, I replaced library normalization plates and Nextera XT tagment Amplicon plates with conventional PCR tubes which may represent a cost-effective alternative. In addition, I recommended the use of equal DNA concentrations of each library during library normalization to ensure better coverage and minimize bias.

2.2 Materials and Method

2.2.1 Bacterial DNA Extraction

For next generation sequencing, contaminant-free, high-molecular weight DNA with an absorbance (260nm/280nm) ratio (1.8 to 2.0) is considered to be a high-quality template DNA (Healey et al., 2014). For DNA extraction, bacteria grown in liquid culture media were pelleted (8,000 g for 8 minutes) in a sterile microfuge tube. The pellet was resuspended in 600 µL phosphate-buffered saline (PBS) until the absorbance at 600 nm (A600) was between 1.0 and 2.0. The cell suspension was lysed by adding 30 µL lysozyme (50 mg/mL) (Muramidase, VWR Chemicals, Radnor, PA, USA), vortexed, and incubated at 37 ºC for 1 hour. The DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) Quick-start protocol with modifications was then used for the extraction of bacterial DNA from the lysed cell suspension. To 200 µL lysate, 20 µL proteinase K and 200 µL buffer AL was added and

mixed by vortexing. 200 µL 100% ethanol was added to the suspension and vortexed. The 620 µL mixture was then transferred to a DNeasy mini spin column placed in a 2mL collection tube and centrifuged at 8,000 g for 1 minute. The spin column was placed in a new collection tube, 500 µL buffer AW1 added, and centrifuged for 1 minute at 8,000 g. The spin column was again placed in a new collection tube, 500 µL buffer AW2 added, and centrifuged at 20,000 g for 3 minutes. The spin column was finally transferred to a 1.5mL sterile microfuge tube and added with 100 µL buffer AE, incubated for 1 minute at room temperature, and centrifuged at 8,000 g for 1 minute to obtain bacterial DNA. The eluted DNA (100 µL) was treated with 2 µL RNAse (100 mg/mL) (Qiagen, Hilden, Germany) and incubated at room temperature for 1 hour. The High Pure PCR Template Preparation Kit (Roche, Basel, Switzerland) was used to purify the RNAse-treated DNA. A modification was made to the manufacturer's protocol in which I performed only four DNA spin-wash steps instead of 9 recommended steps. Initially, 100 µL of binding buffer was added to RNAse treated DNA and incubated at 70 ºC for 10 minutes and then mixed with 50 µL of isopropanol. The content was transferred to a Roche spin column and spun at 8,000g for 1 minute. The flow through was discarded and the spin column was inserted into a new collection tube. A single wash step was performed by adding 500µL wash buffer and spun at 8,000 g for 1 minute. The flow through was discarded and the spin column was inserted into a new collection tube. A final spin (8,000 g for 1 minute) was performed to get rid of excess wash buffer from the spin column. Finally, the spin column was inserted into a 1.5mL sterile microfuge tube and 50µL of pre-heated elution buffer (70 ºC)

were added. The tube was spun at 8,000 g for 1 minute to elute the purified DNA for next generation sequencing.

2.2.2 Quantification of Bacterial Genomic DNA

The use of an accurate concentration of DNA is crucial for bacterial DNA genomic library preparation. To standardize the concentration of DNA, I used the Qubit dsDNA (double-stranded DNA) HS (high sensitivity) Quantification Kit (Life Technologies, Carlsbad, CA, USA) to quantify the amount of DNA. For Qubit assay, 10 µL of standards and 2 µL of sample were added in separate Qubit assay tubes (Life technologies, USA) containing 190 µL and 198 µL of pico green working solution respectively. The Qubit reading was taken after vortexing the mixture for 3 seconds followed by its incubation at room temperature for 2 minutes. The DNA concentration of each sample was adjusted to 0.2 ng/µL after dilution with a required volume of ultrapure™ DNAse/RNAse free distilled water (Invitrogen).

2.2.3 Tagmentation and PCR Amplification of Bacterial Genomic DNA

At this stage, the input DNA is simultaneously fragmented and adapter sequences are added for the subsequent PCR amplification step. This section provides details of the processes involved to fragment and add adapter sequences to template DNA using single tube Nextera XT Tagmentation and to generate multiplexed sequencing libraries. I followed the Nextera XT DNA sample preparation guide (Illumina, 2012) with in-

house modifications in terms of consumables and the volume of specific sample/reagents. I replaced the use of 96-Ill TYC plate with 0.2 mL thin wall clear, flat capped PCR tubes (Fisher scientific, Australia) throughout the method. I also replaced the multichannel pipettes and the high-speed micro plate shaker with single channel pipettes and a bench top centrifuge, respectively.

2.2.3.1 Nextera XT Tagment Amplicon (NTA) Construction

To 2.5 µL (0.2 ng/µL) input DNA in a PCR tube, I added 5 µL tagmentation DNA buffer and 2.5 µL amplification tagmentation mix. The content was briefly vortexed and then transferred to the Applied Biosystems® Veriti 96-Ill thermal cycler (Thermo Fisher Scientific, Foster City, CA, USA) and programmed for a step at 55 ºC for 5 minutes with heated lid, followed by a hold at 10 ºC for a volume of 10 µL.

2.2.3.2 Neutralization of NTA

Immediately after reaching the hold temperature of 10 ºC in the above step, NTA was neutralized by the addition of 2.5 µL neutralization tagmentation buffer and incubation at room temperature for 5 minutes.

2.2.3.3 PCR Amplification

In this stage, the tagmented, neutralized DNA obtained above was amplified using index primers 1 and 2 (Nextera® XT index kit) for cluster formation.

For amplification, 7.5 µL Nextera® PCR master mix and 2.5 µL of index primers 1 and 2 each were added to the tube containing neutralized NTA. The index primers were carefully selected, and labels noted for their use in the later stage of sample sheet preparation during MiSeq WGS. Primer combinations, S502 with N705 /706 and S503 with N701/702 were avoided due to the absence of certain channel colours during whole-genome sequencing (Illumina, 2012). The content was gently pipetted and quickly spun before amplification step which was performed with a Applied Biosystems® Veriti 96-well thermal cycler (Thermo Fisher Scientific, USA) programmed for a working volume of 25 µL with the following settings: heated lid, initial cycle at 72 ºC for 3 minutes followed by 95 ºC for 30 seconds and 12 cycles of (95 ºC for 10 seconds, 55 ºC for 30 seconds and 72 ºC for 30 seconds) with a final run at 72 ºC for 5 minutes followed by a hold temperature of 10 ºC. The amplified, tagmented library was stored at 2 to 4 ºC overnight for PCR clean-up the next day.

2.2.4 Cleaning up the PCR Product

This stage involves cleaning up the amplified PCR product (DNA library) using Agencourt® AMPure XP beads (Beckman Coulter, Brea, CA, USA). The use of magnetic beads allows the selection of appropriately sized library fragments and the removal of very short fragments. For PCR product clean up, AMPure XP beads were brought to room temperature (for 20 minutes) and 80% (v/v) ethanol and 0.2 molar NaOH (40mg NaOH in 5mL distilled water) prepared. To 22.5 µL of PCR product, 11.25 µL of vortexed (30

seconds) AMPure XP beads were added and mixed by pipetting (10 times). The recommended ratio of PCR product and AMPure XP beads is 1:1.6 but I modified the ratio to 1:2. I found this increase in total volume allowed the beads to reach the magnetic ring without affecting the recovery. After incubation at room temperature for 5 minutes, the tubes were placed on a magnetic stand for 2 minutes. While leaving the PCR tubes on the magnetic stand, the supernatant was carefully aspirated and discarded (without aspirating the beads). To the tubes, 100 µL of 80% ethanol was added (beads were not resuspended) and left to stand in the magnetic stand for 30 seconds and the supernatant carefully removed. 100 µL of 80% ethanol was added again. After removing all of the ethanol by pipetting, the tubes containing the beads were removed from the magnetic stand and allowed to air dry in a tube stand for approximately 5 minutes while visually checking for cracks as over drying the beads will significantly reduce elution efficiency. Once the tubes were dried, 26.15 µL of resuspension buffer was added with gentle pipetting 20 times to mix. The tubes were incubated at room temperature for 2 minutes and then placed again on a magnetic stand for 2 minutes (until the supernatant cleared), followed by transfer of the supernatant to new PCR tubes. Although the final supernatant could be stored at -15 to -20°C for up to 1 week, I recommend proceeding to library normalization immediately.

2.2.5 Library Normalization

The cleaned-up PCR product obtained from different samples may not contain an equal concentration of DNA. A pool of accurate input DNA quantity results in more even genome coverage with minimal bias and ensures consistent and reproducible results downstream (Doyle, 2015). For minimally concentrated DNA template, an increased number of PCR cycles to obtain a high amount of PCR product is contraindicated as it impairs sequence heterogeneity and leads to poorer sequence quality (Robin et al., 2016). In contrast, for a sample with a high amount of PCR product, dilution is recommended as a decreased amount of template DNA may increase the ratio of duplicate reads and also increase coverage in low GC regions (Parkinson et al., 2012). I, therefore, prepared a pool of DNA from different samples such that the concentration of each was represented equally for the sequencing run.

The Qubit DNA quantification method as described previously was used to determine library concentration in each sample. The sample with the lowest DNA concentration was used in a volume of 10µL as the reference to prepare a library pool using the formula:

Volume required (V2) = Concentration original (S1) x Volume total (V1 = 10 µL) divided by Concentration required (S2).

In this modified step, I normalised the library by using LNS1 (Library Normalisation Storage Buffer 1) only. To X µL of library pool, X µL of freshly prepared 0.2molar NaOH (final concentration 0.1 molar) was added and incubated for 5 minutes at room temperature. To the NaOH treated

suspension, an equal volume (2X µL) of LNS1 was added from the Nextera® XT library preparation kit. From this, a 1:1000 dilution of (1 µL of sample added to 1000 µL of ultrapure distilled water) was prepared for the quantitative PCR (qPCR) run.

2.2.6 Quantification of the NaOH Treated Pooled Library

To check the concentration of pooled library, I used KAPA library quantification kit Illumina® (No ROX) for quantitative PCR on a Corbett Rotor-Gene 6000 real-time thermocycler (Corbett Research, Rotor-gene) using the following set up: a hot start run at 95ºC for 10 minutes followed by 40 cycles of (95ºC for 10 seconds and 60ºC for 30 seconds). A set of six DNA standards (with concentrations ranging from 20 picomolar (pM) to 0.0002 pM), three sets of negative control (ultrapure distilled water), and three sets of the DNA library were included in the qPCR run.

Following the completion of the qPCR, the concentration of DNA in the pooled library was determined by the standard curve method and calculated in pM concentration per tube. To calculate the original concentration of the pooled library I applied the formula:

Average sample concentration (in pM) * insert size standards (452 base pairs) * dilution factor (1000) divided by insert size of the pooled library (500 base pairs)

The concentration and size profile of the library is dependent on the amount of input DNA. The average fragment size ranged from 250 base pairs to 1.5

Kb in length and an average library fragment size of 500 base pairs was used for the calculation (Illumina, 2014).

This value obtained from the calculation represented the concentration of DNA in the pooled library and was used to estimate the dilution factor required to achieve a final library concentration of 15 pM in a 600 µL volume (using buffer HT as a diluent).

For example, for a qPCR determined concentration of 2.36 pM in a 1:1000 dilution of the pooled library, the library DNA concentration is determined as follows

(2.36 pM * 452 base pairs * 1000) /500 base pairs = 2135.25 pM (library concentration)

To calculate the volume of the pooled library to dispense onto the MiSeq reagent cartridge, the following formula is used:

Volume required = (Concentration required * Volume total) divided by Concentration original

i.e. = (15pM * 600 µL) / 2135.25 pM

= 4.22 µL

Therefore, 4.22 µL is added to 595.78 µL of HT buffer to produce a final concentration of 15 pM, in a final volume of 600 µL. The diluted library is then ready to be heat denatured and loaded into the MiSeq reagent cartridge.

2.2.7 Preparing Pooled Library for Loading onto MiSeq

The pooled amplified library (PAL) was thawed at room temperature and mixed by pipetting up and down (5 times) and briefly centrifuged. Based on the library concentration example above, I then transferred 595.78 µL of HT buffer to a 1.5mL diluted amplified library (DAL) tube containing 4.22 µL PAL and mixed using a pipette (5 times). The DAL tube was then vortexed at top speed, centrifuged briefly, and incubated exactly for 2 minutes at 96 ± 2 ºC. The DAL tube was then immediately transferred to ice for at least 5 minutes. Prior to commencing the whole-genome sequencing process, the Illumina MiSeq sequencer was put through a short wash cycle to avoid cross-contamination of the DAL from the previous usage. According to the manufacturer's protocol, the MiSeq reagent cartridge was thawed at room temperature. The Illumina Experiment Manager was used to generate a MiSeq sample sheet with the following configuration:

Generate FASTQ workflow; FASTQ Only application; NexteraXT assay; 151 insert reads; assignment of the samples with a unique identifier and index-pair combination.

The entire 600 µL of DAL was finally loaded onto the MiSeq reagent cartridge. Following the setup procedure of the Illumina Experiment Manager in which the flow cell was rinsed with MilliQ water and soak dried, the reagent buffer and the DAL loaded MiSeq reagent cartridge was inserted into MiSeq instrument for sequencing to commence. The raw Fastq sequence reads from whole-genome sequencing were stored on the local

computer as well as on the Illumina BaseSpace server (https://basespace.illumina.com/) for further analysis.

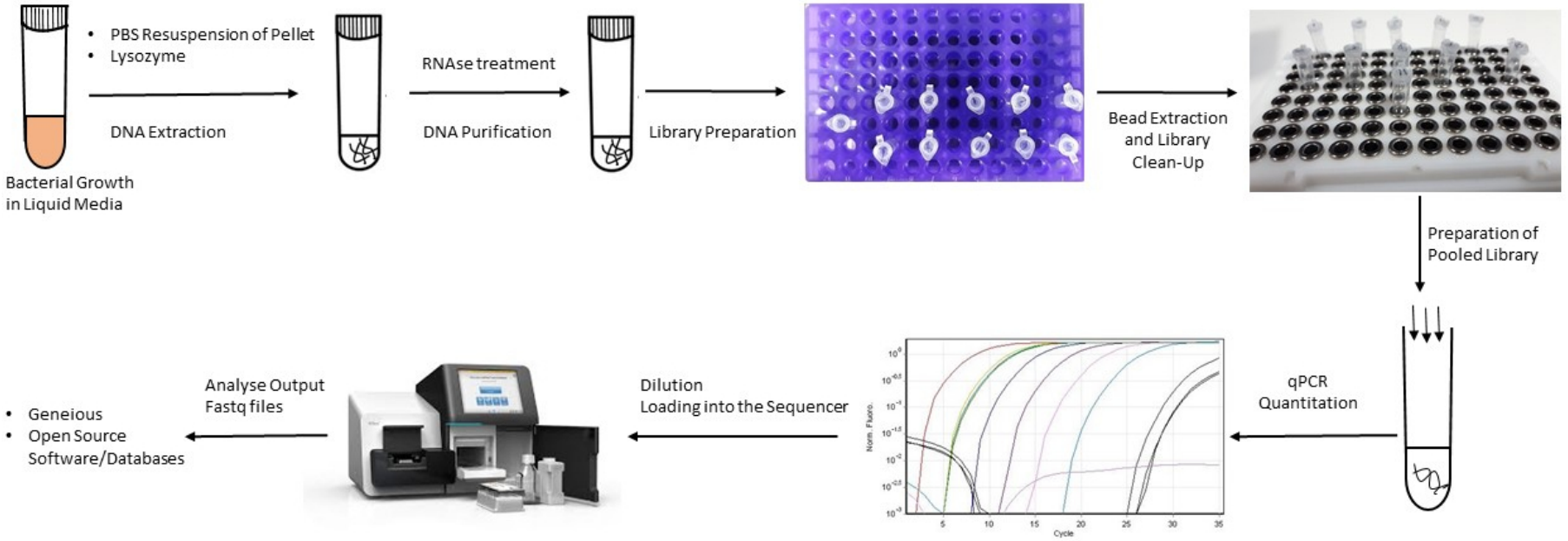


Figure 2.1. Schematic diagram showing processes involved in generating whole-genome sequence data of *M. tuberculosis.*

2.3 Bioinformatics Analysis

2.3.1 Data Analyses Tools

The selection of bioinformatics software for the analysis of WGS data may be determined according to the objective of the study. Here, I used Geneious 9.1.8 (Biomatters Ltd.) (Kearse et al., 2012), a desktop software to map, find variants, locate differentially encoded genes and build a phylogenetic tree using the available plugins (Supplement 1). In addition, I used Galaxy Australia (Afgan et al., 2015) based tools, for example, FastQC (Andrews, 2010), Burrows-Wheeler Aligner (Li & Durbin, 2010), SAM-tools and SnpEff (Li et al., 2009) to examine the sequenced data. The data were also analysed and by a collaborator at Trinity College Dublin, Ireland. A consensus was made by comparing the results obtained by myself and the collaborator before reaching to a final conclusion and submission for the publications. The open source databases, for example, TGS-TB (Sekizuka et al., 2015), PhyResSe (Feuerriegel et al., 2015), COG (Tatusov et al., 2000), BLAST (Altschul et al., 1997) and KEGG (Kanehisa & Goto, 2000) were used to further analyse the whole-genome sequence data of tubercle bacilli. These databases enabled the acquisition of information on bacterial pathogens that include genotype and phylogeny and antibiotic-resistance mutations. For example, *M. tuberculosis* complex raw fastq.gz files uploaded to the TGS-TB database (Sekizuka et al., 2015) produced results in terms of presence of mutations determining drug resistance, *in silico* spoligotyping, identification of lineage type, and phylogenetic tree

building with respect to the available strains in the database. This database also enabled the detection of IS*6110* insertion sites, and 43 customized loci for variable number tandem repeat (VNTR). A list of bioinformatics tools and databases used for the analyses of whole-genome sequence data in the thesis is summarised in table 3.

This method of generating whole-genome sequence data of bacteria including *M. tuberculosis* was based on the manufacturer's instructions (Illumina, Inc. 2012). Modifications were done in terms of the usage of consumables and library normalization step to maximize the quality of output data (Gautam et al., 2019a). The proposed method was not compared for its sensitivity and accuracy with any published protocol; however, we were able to reproduce the protocol for studies involving *M. tuberculosis* and other bacterial pathogens (Gautam et al., 2019; Gautam et al., 2018; Gautam, Mac Aogáin, & O'Toole, 2017; Leong et al., 2018; Mac Aogáin et al., 2016).

Table 3. Bioinformatics tools and databases used across different chapters in the present thesis

Databases/Tools	Purpose	Chapters
Burrows-Wheeler aligner	Mapping	3,4,5
SAMtools	Variant detection	3,4,5
SnpEff	Variant annotation	3,4,5
SPAdes	de novo assembly of contigs	3,4,5
PhyResSE	Drug-resistant mutation detection	3,4,5
Geneious	Trimming, mapping, variant detection, *de novo* assembly, contigs and consensus sequence generation, open reading frame analyses, primer design, phylogenetic tree construction	3,4,5
BLASTp	Search protein databases	3
BLASTn	Search nucleotide databases	3
KEGG	Identification of a cluster of orthologous groups	3
PhymL	Phylogenetic tree construction	3,4
TB Profiler	Lineage prediction and detection of drug resistant conferring polymorphisms	4,5
TGS-TB	*In silico* spoligotyping	4
ABACAS	Order assembled contigs	5
Galaxy	Determine the quality of sequenced data using FastQC	3,4,5

Chapter 3

Investigating Virulence Related Genes in an Outbreak Strain of *Mycobacterium tuberculosis* using Whole-Genome Sequencing

3.1 Tuberculosis Outbreak

Tuberculosis (TB) accounts for approximately five thousand deaths per day. The majority (55%) of the total new TB cases in 2017 were reported across five countries including India, China, Indonesia, the Philippines and Pakistan (in descending order) (World Health Organization, 2018d). However, outbreaks of TB can occur in low as well as high TB burden countries. For example, *M. tuberculosis* strain CDC1551 accounted for an outbreak of TB during 1994-1996 in two low TB burden (<1 case per year) counties (Tennessee and Kentucky) of the USA where 311 out of 429 contacts of TB patients were detected positive for *M. tuberculosis* infection in a skin test (Valway et al., 1998). Similarly, an outbreak of TB was identified in a high incidence (2901 per 100,000 population) Daru island of Papua New Guinea where 0.76% of total 16,714 population were diagnosed with multi-drug resistant TB (Bainomugisa et al., 2018). The factors associated with increased TB transmission may be categorised under microbial properties and the hosts' environmental circumstances. For example, the CDC1551 strain is associated with increased virulence (Valway et al., 1998) and *M. tuberculosis* Beijing strain is related to their higher transmission rate (Iwamoto et al., 2009; Yang et al., 2012). The host environmental factors favouring transmission include household

overcrowding (Waziri et al., 2014), homelessness, co-infection with HIV, increased median infectious period (Powell et al., 2017), mental illness and substance abuse (Dobbins et al., 2012).

Although a prolonged period of latency after *M. tuberculosis* exposure may affect the reliability to describe an event of TB outbreak, the definition for TB outbreak been established to prioritise the resource allocation in such circumstances (Denholm et al., 2016). The description of TB outbreak varies with the local situation. For example, the CDC criteria to identify a TB outbreak is primarily based on the surveillance data where notifications of a higher number of TB cases than predicted, detection of TB in two or more connected cases during contact investigation, an established link between two or more cases diagnosed within a year that may not have been identified during contact investigation and genotypic evidence of transmission that was not detected during contact investigation during the preceding two years is flagged as an outbreak of TB (Freimanis e al., 2007). Furthermore, the outbreak is confirmed when control measures are unable to disrupt the ongoing TB transmission and when TB investigation programs require additional assistance (Freimanis et al., 2007). The Communicable Diseases Intelligence (CDI) publication of the Australian Department of Health provides definitions for a cluster of TB cases, a probable cluster, a possible cluster, and an outbreak of TB (Denholm et al., 2016). Any two or more active cases of TB with known epidemiological links and related to the same genotype of *M. tuberculosis* are defined as TB cluster cases (Denholm et al., 2016). The outbreak of TB is described when active cases of TB are detected in three or more members (where at least two members transmit

disease) with an epidemiological and microbial genotypic linkage supporting evidence of serial transmission (Denholm et al., 2016).

3.2 Investigating the Genome of an Outbreak Strain of *M. tuberculosis* in New Zealand

3.2.1 Tuberculosis in New Zealand

New Zealand is a low TB incidence country with 310 cases reported among 4.7 million people in 2017 (World Health Organization, 2018h). The majority of TB cases in New Zealand have been detected among individuals belonging to Middle Eastern/Latin American/African ethnic groups (22.1 per 100,000 population) and residents of socio-economically disadvantaged locations (Institute of Environmental Science and Research Ltd, 2015). Individuals born overseas or living with people not born in New Zealand were the most common risk factors associated with TB (Institute of Environmental Science and Research Ltd, 2015). A discrete distribution of cases was noted across different district health boards in New Zealand where the national rate of TB incidence is significantly affected by events of TB outbreak. For example, two outbreaks of TB added 14 new cases to the national notification data in 2014 (Institute of Environmental Science and Research Ltd, 2015).

3.2.2 *Mycobacterium tuberculosis* "Rangipo" Strain Outbreak in New Zealand

An analysis of a TB outbreak that occurred between 1996 and mid 2000 in New Zealand identified a mother of the newly born child as an index case (De Zoysa et al., 2001). Based on the IS*6110* restriction-fragment length polymorphism DNA typing, 43 out of 61 TB cases were identified as members of the same TB outbreak that occurred across eight districts over a period of three and half years (De Zoysa et al., 2001). The remaining 18 cases were epidemiologically related to the outbreak (De Zoysa et al., 2001). A majority (86.9%) of these cases were detected in the indigenous Māori population (De Zoysa et al., 2001). One of the patients related to the outbreak served his sentence in the Tongariro/Rangipo prison in 1998 which led to the derivation of the strains name, "Rangipo" (McElnay, Thornley, & Armstrong, 2004). In 2002, an outbreak of TB due to Rangipo strain involving 19 active cases was identified in the Hawke's Bay area on New Zealand's North Island (McElnay, Thornley, & Armstrong, 2004). The high transmissibility of the Rangipo strain was demonstrated by the detection of active TB in 16.4% and latent TB in 20% of close contacts (McElnay, Thornley, & Armstrong, 2004). This particular strain constituted a major proportion of TB cases in 13 out of 18 district health boards between 2003 and 2007 in New Zealand (Sexton, Petera, & Pandey, 2008). Based on the large sequence polymorphism and SNP analyses, the Rangipo strain is phylogenetically classified as the Euro-American lineage 4 *M. tuberculosis* (Yen et al., 2013). Although the high infective potential of the Rangipo strain

has been proposed previously (McElnay, Thornley, & Armstrong, 2004), experimental evidence to support this claim is limited.

In this study, I performed whole-genome sequencing of *M. tuberculosis* Rangipo strain to explore the virulence related properties encoded in its genome. This chapter is published online (Gautam et al., 2017a) and can be accessed here: https://doi.org/10.1080/23744235.2017.1330553

3.2.3 Materials and Methods

3.2.3.1 Culture and MIRU Typing of the Rangipo Strains

The Rangipo isolates were grown in Mycobacteria Growth Indicator Tube (MGIT) media at LabPLUS Limited, Auckland City Hospital and Mycobacterial Interspersed Repetitive Unit (MIRU) 24-loci typing was performed as described previously (Cole et al., 1998). The MIRU clusters were confirmed when the isolates possessed the same number of copies across all the loci. Eight isolates displayed a common 24-loci MIRU profile (233325153324 341444223362) while the ninth isolate (sample 356) varied in one of the loci. Whole-genome sequencing was performed on purified DNA extracts from nine Rangipo isolates as explained in Chapter 2 generating paired-end Fastq reads.

3.2.3.2 Bioinformatics Data Analyses

The Fastq sequence data files were analysed using the Geneious R9.0 software suite (Kearse et al., 2012). The reads were trimmed with an error probability limit set at 0.05 to remove the Illumina adapter sequences.

Initially, the sequences were mapped against the reference strain H37Rv (accession number NC_000962.3), a publicly available and widely used annotated reference genome of *M. tuberculosis* (Cole et al., 1998). For mapping, I used random multiple base match set at a maximum variant *p*-value of 10^{-6} when exceeding 65% bias. Single-locus variations (SLVs) and insertions and deletions (indels) were identified using the BWA/SAMtools mapping approach as described previously (Mac Aogain et al., 2016). Briefly, I analysed regions with a minimum read depth of 20 and a minimum variant frequency of 95%. This allowed us to detect SLVs in Rangipo strains with respect to the reference genome, H37Rv. The genes containing non-synonymous SLVs were categorised into the different cluster of orthologous groups (COGs) based on their known or predicted function. In addition, the literature was searched to relate the role of these mutations in *M. tuberculosis* virulence (Forrellad et al., 2013).

After analysing the SLVs, I was interested to examine if the Rangipo strain possessed additional genes that are related to the physiology or pathogenesis of *M. tuberculosis*. It is apparent that not all the sequences in the test strain align with the reference genome during the process of genome assembly. To identify genes that were present in the Rangipo strain but absent in the *M. tuberculosis* reference strain H37Rv, I took the unmapped reads obtained after mapping with H37Rv. These unmapped reads were de-novo assembled to generate a consensus sequence. From the consensus sequence, open reading frames (ORFs) of at least 150 nucleotides with start codons ATG, TTG or CTG and standard genetic code were produced. I selected interior ORFs for the analyses assuming that start

and/or stop codons may be positioned away from the minimum length of ORF. The analyses of ORFs in terms of their predicted protein product present in *M. tuberculosis* was done using the BLASTP database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990) where the search was targeted for non-redundant protein sequences. This represented gene insertions in Rangipo strain relative to H37Rv. Furthermore, to identify the orthologues of additional copies of Rangipo genes (that were absent in H37Rv) in other strains of *M. tuberculosis*, a cut-off of ≥ 98 % in terms of subject and query sequence match was used.

To infer the phylogenetic relationship between the isolates, the consensus sequence of each Rangipo genome generated after mapping with *M. tuberculosis* reference strain H37Rv was used. The generalised time reversible (GTR) substitution model was used to derive a maximum likelihood (ML) phylogenetic tree using PhyML where 10 starting trees were selected and bootstrapping was performed over 100 replicates (Guindon & Gascuel, 2003). The PhyML ML algorithm was used to infer the phylogenetic relationship between the isolates as it allows us to input alignment of nucleotides for the analyses and has been widely used as a reliable method to test the biological hypothesis (Guindon et al., 2009; Casali et al., 2014). In addition to the nine Rangipo strains, four annotated whole-genomes of lineage 4 *M. tuberculosis* strains, H37Rv (NC_000962), CDC1551 (NC_002755), Haarlem (NC_022350), and Erdman (NC_020559) were included in the phylogeny construction.

3.2.3.3 Description of the Whole-Genome Features of a Rangipo Strain and Deposition of the Sequence in Public Databases

A representative isolate of Rangipo strain (isolate number 494) was mapped to the *M. tuberculosis* strain H37Rv reference genome (accession no. NC_000962) using the Burrows-Wheeler Aligner (Li & Durbin, 2010). The contigs obtained after mapping were de novo assembled using the SPAdes assembler (version 3.7) (Bankevich et al., 2012). The number of variants in the Rangipo genome with respect to the H37Rv reference genome was obtained using the SAM-tools platform. These variants were then annotated using the SnpEff analyses suite (Cingolani et al., 2012). To identify the presence of drug resistance determining mutations in the Rangipo strain, paired-end raw Fastq files were imported into the publicly available PhyResSE database (Feuerriegel et al., 2015). The whole-genome shotgun project of an isolate 494 was submitted for deposition at DNA Data Bank of Japan (DDBJ), European Nucleotide Archive (ENA) of European Molecular Biology Laboratory (EMBL), and the National Institute of Health's genetic sequence database, GenBank and is obtained under accession number LXWG00000000.

3.2.4 Results

3.2.4.1 Single Locus Variations in the TB Outbreak 'Rangipo' Strain

A total of 727 single locus variations (SLVs) with respect to the reference H37Rv strain were identified across all the nine Rangipo genomes. Of these, 700 SLVs were common to all of the Rangipo isolates sequenced.

The result is comparable to that of Colangeli and colleagues who reported the presence of 747 single nucleotide polymorphisms (SNPs) with respect to H37Rv in 10 Rangipo isolates studied (Colangeli et al., 2014). The Rangipo common SLVs in this study constituted 386 non-synonymous SLVs which were further categorised as missense variants (n=354), frameshift mutations (n=21), in-frame insertions (n=4), in-frame deletions (n=2), gains in stop codon (n=4) and loss in stop codon (n=1). To identify TB cases belonging to the same transmission network, a cut-off of ≤ 5 SNP differences between the isolates was used as described previously (Nikolayevskyy, Niemann, & Drobniewski, 2016; Walker et al., 2013).

The number of SNP differences between isolates ranged between 0-19. For example, isolates 431 and 356 exhibited no SNP variation while isolates 547 and 278 were separated by 19 SNPs (Table 4) (Gautam et al., 2017a).

Table 4. Single-locus variation differences between the nine Rangipo isolates mapped to the *M. tuberculosis* reference genome, H37Rv.

	376	278	431	356	547	22	494	486	467
376		17	8	8	10	6	6	7	7
278			11	11	19	15	15	16	16
431				0	10	6	6	7	7
356					10	6	6	7	7
547						8	8	9	9
22							4	5	5
494								1	1
486									0
467									

Reproduced with permission from Gautam et al. Infect Dis (Lond). (2017). Copyright Informa UK Limited.

3.2.4.2 Functional Classification of SLVs Detected in the Rangipo Strain

Across the 700 Rangipo common SLVs, four coding sequences including Rv0872c, Rv1435c, Rv2407 and Rv2823c had in-frame insertions whereas in-frame deletions were identified in two coding sequences, Rv3345c and Rv0849. Similarly, 4 genes (*pstA1* (Rv0930), *ipdA* (Rv3303), PE35 (Rv3872) and Rv0851c) gained a stop codon while the polyketide beta-ketoacyl synthase *pks3* gene (Rv1180) lost a stop codon. In addition, 55 variants were identified in genes encoding members of the PPE and PE_PGRS families of proteins. Non-synonymous single locus variants, common to all Rangipo genomes sequenced, were detected in 12 genes that have been linked to the virulence of *M. tuberculosis* (Forrellad et al., 2013). These included SLVs in *aceAa, fadD28, kefB, mce1F, mycP1, pckA, pepD, phoR, pks15, plcB, pstA1* and *pstS1* (Table 5) (Gautam et al., 2017a).

Table 5. Non-synonymous single-locus variations detected in the nine Rangipo isolates sequenced in genes related to mycobacterial virulence.

Locus tag	Gene	Role of the gene in virulence	Nucleotide change	Amino acid change
Rv0174	*mce1F*	Growth regulation in murine macrophages	1109T>C	Leu370Pro
Rv0211	*pckA*	Growth regulation in the human monocyte cell line	302A>C	Asn101Thr
Rv0758	*phoR*	Regulation of cell wall hydrophobicity	515C>T	Pro172Leu
Rv0930	*pstA1*	Resistance to host immunity	913C>T	Arg305[a]
Rv0934	*pstS1*	Adhesin binding to macrophages	63_64insA	Ala22[b]
Rv0983	*pepD*	Stress response protein	1169T>C	Leu390Pro
Rv1915	*aceAa*	Growth regulation in acetate containing medium	26A>C	Glu9Ala
Rv2350c	*plcB*	Growth regulation in macrophages	1406T>C	Leu469Ser
Rv2941	*fadD28*	Virulence attenuation in BALB/c mice	545T>C	Val182Ala
Rv2947c	*pks15*	Interaction with host cells	998T>C	Val333Ala
Rv3236c	*kefB*	Bacterial persistence in Guinea pig	962G>T	Arg321Leu
Rv3883c	*mycP1*	Regulation of ESX-1	979C>T	Pro327Ser

[a]Stop gained. [b]Frameshift variant.

3.2.4.3 Classification of Genes Containing Non-Synonymous Variants into Different Cluster of Orthologous Group Categories

Genes containing non-synonymous variants in the Rangipo strain were classified into functional classes of the cluster of orthologous group (COG) (Galperin et al., 2015). A total of 411 NS-SLVs were analysed for their distribution into different COG categories. This resulted in the classification of 156 genes into single COG categories and 11 genes into two COG categories with the remaining 219 genes unassigned based on the current KEGG database (http://www.genome.jp/kegg/) (Kanehisa et al., 2016). A higher proportion (59.5%) of NS-SLVs present in all Rangipo strains were detected in genes related to metabolic function categorised as COG groups C, E, F, G, H, P, Q, R, S, and I. This was followed by categories D, M, N, O, T and V (23.5%) associated with cellular process and signalling. In contrast, the lowest numbers of genes were classified under nucleotide transport and metabolism COG F (n = 1) and cell motility COG N (n = 2) groups (Table 6) (Gautam et al., 2017a). The functional impact of the number of polymorphisms in each COG category has not been widely reported. However, we can speculate that the higher number of polymorphisms in genes related to metabolic function in Rangipo strain may be related to the production of mutant enzyme required for their adaptation in the surroundings (Ilina et al., 2013). Furthermore, a low number of NS-SLVs related to *M. tuberculosis* cellular process and signaling suggests the evolutionary pressure in the genes of this category (Ilina et al., 2013). Fewer numbers of NS-SLVs in COG N and COG V may indicate conserved cell surface structure of the Rangipo strain and its non- extremely drug resistant

phenotype respectively (Ilina et al., 2013). Notably, the number of genes in each category should be carefully interpreted as their proportion in the COG system more likely impacts the observation in each category. For example, a greater number of genes assigned in the C, H and L (102, 123, 81, respectively) categories of the COG system.

Table 6. Classification of Rangipo common genes containing non-synonymous single-locus variations into different clusters of orthologous groups (COGs).

COG	Functional category	Gene (n)
C	Energy production and conversion	16
D	Cell cycle control, cell division, chromosome partitioning	3
E	Amino acid transport and metabolism	12
F	Nucleotide transport and metabolism	1
G	Carbohydrate transport and metabolism	9
H	Coenzyme transport and metabolism	14
I	Lipid transport and metabolism	11
J	Translation, ribosomal structure and biogenesis	11
K	Transcription	4
L	Replication, recombination and repair	15
M	Cell wall/membrane/envelope biogenesis	13
N	Cell motility	2
O	Post translational modification, protein turnover, chaperones	9
P	Inorganic ion transport and metabolism	14
Q	Secondary metabolite biosynthesis, transport and catabolism	8
R	General function prediction only	9
S	Function unknown	12
T	Signal transduction mechanism	8
V	Defence mechanism	7
	COG not designated	219

3.2.4.3 Identification of Differential Gene content in the 'Rangipo' Strain

To determine whether the Rangipo strain carries genes that are absent in the reference H37Rv strain, I investigated the unmapped reads (generated after assembly with H37Rv) that otherwise, in routine analyses would have gone unnoticed. This resulted in the detection of five genes in the Rangipo strain that were not present in the H37Rv genome. Upon analysing the presence of these genes in other *M. tuberculosis* strains using BLASTP analyses (Altschul et al., 1990), an orthologue of each gene was detected in *M. tuberculosis* CDC1551 strain. These include orthologues of transcriptional regulator protein EmbR2 (MT3428), molybdopterin cofactor biosynthesis protein A (MT3427), molybdopterin cofactor biosynthesis protein B (MT3426), an extended-length DNA helicase (MT2082), and an additional copy of adenylate cyclase (MT1360) in CDC1551 strain. Polymerase chain reaction rather than Southern Blotting was performed on the Rangipo isolates and H37Rv using primers for the genes (MT3426, MT3427, MT3428, MT2082 and MT1360) to confirm their presence and absence as indicated by the genome sequence analysis. No amplified products were observed in H37Rv in the agarose gel electrophoresis. However, bands were visible for all the five genes in the Rangipo strain which were confirmed by the Sanger sequencing of PCR amplified products. Presence of these five genes has previously been reported in the annotated genome of *M. tuberculosis* strain CDC1551 by Fleischmann et al (Fleischmann et al., 2002). Furthermore, the Rangipo associated five genes were examined in the annotated genomes of additional *M. tuberculosis*

lineage 4 strains i.e. Erdman and Haarlem. The Erdman strain possessed an orthologue gene for MT3427 while Haarlem strain contained orthologues of MT3428, MT3427 and MT2082 (Figure 3.1) (Gautam et al., 2017a). However, a preliminary investigation suggests the presence of sequence homology of Rv3109 (MoaA1) and Rv0869c (MoaA2) in the Rangipo strain with 100% similarity at the nucleotide level.

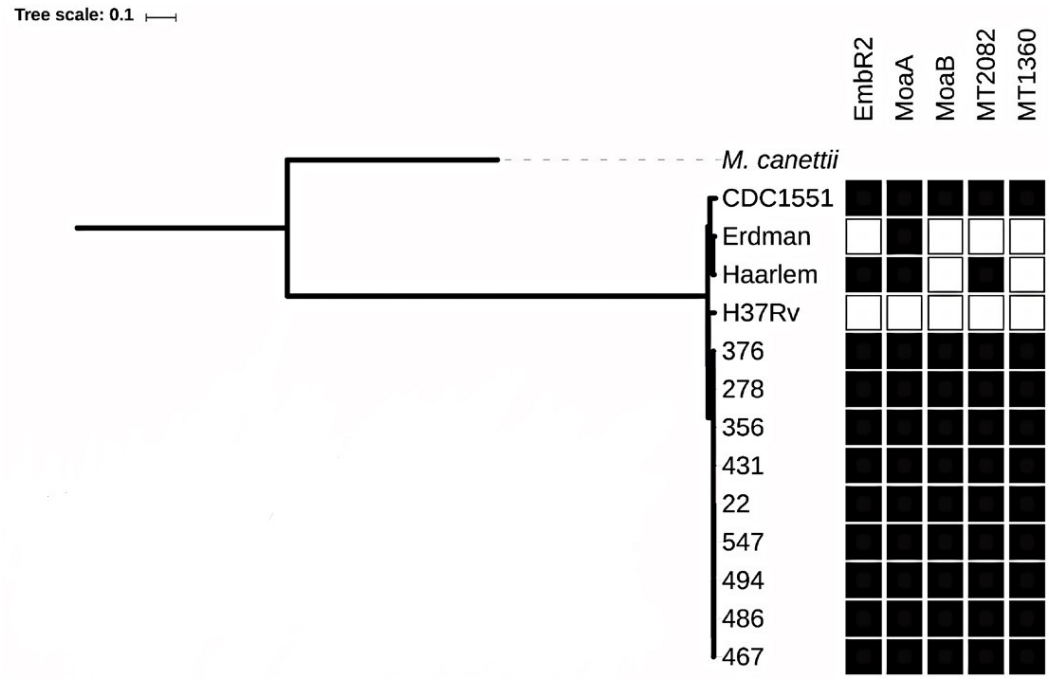


Figure 3.1. Identification of additional genes in the Rangipo strain and annotated reference genomes of *M. tuberculosis*.

A generalized time reversible (GTR) substitution model-based maximum-likelihood phylogenetic tree was built including nine Rangipo strains and four lineage 4 reference strains: CDC1551 (SRX393042); Haarlem (SRX347319); Erdman (SRX364193); and H37Rv (NC_000962.3) using PhyML. *M. canettii* reference genome (HE572590.1) was used to root the

tree. Black squares represent the presence of a particular gene while clear squares indicate their absence.

Reproduced with permission from Gautam et al. Infect Dis (Lond). (2017). Copyright Informa UK Limited.

3.2.4.5 Consensus Genome Assembly

A total of 2,128,439 paired-end reads of Rangipo isolate 494 were mapped to the H37Rv reference genome. This yielded an average read depth of 45-fold, covering 99.5% of the reference genome. A total of 851 variant sites were identified relative to the H37Rv genome and consisted of 782 single-nucleotide variants (SNVs) and 69 insertions/deletions. Five hundred and eighty-seven of the variants were nonsynonymous, of which 549 were SNVs and 38 were insertions/deletions. Drug resistance determining mutations were predicted using PhyResSe database (Feuerriegel et al., 2015) which identifies the mutations associated with first and second-line drug resistance in TB from whole-genome sequence data. For example, the database detects single nucleotide polymorphisms in *katG, inhA, ahpC, rrs, rpsL, embA* and *embC* in 100% agreement with other methods (Feuerriegel et al., 2015). In terms of genotypic drug susceptibility testing, no mutation associated with drug resistance *in M. tuberculosis* was detected in Rangipo isolate 494 which is consistent with the phenotypic drug susceptibility profile as determined on MGIT cultures of the isolate by LabPLUS Ltd, Auckland.

For public access, the whole-genome shotgun project of *M. tuberculosis* Rangipo strain isolate number 494 has been deposited in DDBJ/EMBL/GenBank under the accession no. LXWG00000000.

3.3 Discussion

Multiple outbreaks of TB in New Zealand have been associated with the *M. tuberculosis* Rangipo strain (De Zoysa et al., 2001; McElnay et al., 2004). In this work, I performed whole-genome sequencing to investigate the genome of the Rangipo strain, in particular, variants in virulence-related genes. A total of nine isolates representing the Rangipo genotype were sequenced and analysed with respect to *M. tuberculosis* reference strains, H37Rv, CDC1551, Haarlem and Erdman.

There were 700 single locus variants common to all nine isolates of Rangipo strains when mapped with the genome of the reference strain, H37Rv. The majority of these polymorphisms (n=592, 84.5%) were detected in the coding regions which consisted of 386 non-synonymous variants. Among these 386 NS-SLVs, 12 were detected in genes that were previously linked to mycobacterial virulence (Forellad et al., 2013). These include the genes *aceAa, fadD28, kefB, mce1F, mycP1, pckA, pepD, phoR, pks15, plcB, pstA1* and *pstS1*. For example, decreased number of *M. tuberculosis* in the spleen and lungs of infected mice and their poor survival in macrophage cells derived from murine and human monocytes have been associated with a mutation in *aceAa*, isocitrate lyase gene (Munoz-Elias & McKinney, 2005). An essential role of the *PhoPR* system has been described for *M. tuberculosis* growth under low Mg^{2+} concentrations *in vitro*, and in

macrophage and mouse infection models (Walters et al., 2006). A reduction in phenolic glycolipid production associated with a *pks15/1* mutation has been related to decreased *M. tuberculosis* virulence in the meninges of rabbits (Tsenova et al., 2005). Similarly, triple *plcABC* or quadruple *plcABCD* mutations have been found to exhibit reduced growth of *M. tuberculosis* in the mouse infection model during their late phase of infection (Raynaud et al., 2002). It should be noted that the polymorphisms reported above are present in genes that have been previously related to the virulence of *M. tuberculosis* and may or may not alter the gene function. Therefore, validation of their specific role in the pathogenesis of *M. tuberculosis* requires further experimentation.

In addition to non-synonymous SLVs, this study identified the differential carriage of genes in the Rangipo strain that were absent in H37Rv. The orthologues of additional Rangipo genes were identified in a clinical strain, CDC1551 associated with an outbreak of TB. Three of these genes have previously been related to mycobacterial virulence and include CDC1551 orthologue MT3428 (transcriptional regulator protein *EmbR2*), MT3427 (molybdopterin cofactor biosynthesis protein A), and MT3426 (molybdopterin cofactor biosynthesis protein B). An orthologue of the H37Rv transcriptional regulator *embR* was reported in CDC1551 strain (Molle et al., 2008). The phosphorylation of *embR* by serine/threonine protein kinase (*pknH*) leads to the activation of the *embCAB* gene associated with the biosynthesis of a key *M. tuberculosis* cell wall constituent, arabinogalactan. In contrast to *embR, embR2* is not phosphorylated by *pknH* (Molle et al., 2008) but inhibits the *pknH* autokinase

activity and *embR* phosphoryl transfer (Papavinasasundaram et al., 2005). In comparison to the parental strain of *M. tuberculosis*, a *pknH* mutant replicated and survived more resulting in the increased load of infecting bacilli in the lungs and spleen of the mouse (Papavinasasundaram et al., 2005). Hence, the role of e*mbR2* in mycobacterial physiology and pathogenesis has been predicted through its regulation of the *pknH/embR* pair (Molle et al., 2008). In addition to the Rangipo and CDC1551 strains, *embR2* orthologues were detected in the lineage 4 Haarlem strain but not in the Erdman strain of *M. tuberculosis* (Figure 3.1).

The orthologues of MT3427 and MT3426 in CDC1551 have been related to the production of molybdopterin cofactor (MoCo) biosynthesis proteins. MoCo functions as an essential cofactor for enzymes involved in redox reactions, for example, *narGHI*-encoded nitrate reductase that has been reported to be involved in *M. tuberculosis* adaptation in hypoxic conditions (Williams, Mizrahi, & Kana, 2014) and persistence in guinea pig lungs (Williams et al., 2015). The MT3427, MoCo biosynthesis protein A (MoaA) participates in the conversion of guanosine triphosphate to cyclic pyranopterin monophosphate (Williams, Kana, & Mizrahi, 2011). On the other hand, MT3426, MoCo biosynthesis protein B (MoaB) is involved in a key step of MoCo biosynthesis, the adenylation of molybdopterin (Williams, Mizrahi, & Kana, 2014). In addition to the above mentioned three genes, an extended-length DNA helicase (MT2082, 1606 amino acids), and an additional copy of adenylate cyclase (MT1360) were also identified in the Rangipo strain but not in the reference strain H37Rv. The functional role of

the extended length DNA helicase and the extra copy of adenylate cyclase in the virulence of the Rangipo strain awaits further studies.

An observation similar to the Rangipo and CDC1551 strain in terms of the presence of five additional genes was described in clinical isolates from Russia (Azhikina et al., 2006). In addition to these genes, the authors reported the detection of a truncated copy of *plcD* gene corresponding to the first 843 nucleotides of total 1,545 base pairs of the MT1799 gene in CDC1551. Notably, a truncated *plcD* gene was detected in the Rangipo strain which had first 905 nucleotides similar to CDC 1551 orthologue MT1799. Furthermore, I analysed the publicly available whole-genome sequences of *M. tuberculosis* strains associated with outbreaks of TB in Germany (Roetzer et al., 2013), Guatemala (Saelens et al., 2015), Korea (Han et al., 2015), and the UK (Török et al., 2013) for their presence of five additional genes detected in the New Zealand outbreak Rangipo strain. All the genomes of the outbreak strains investigated were similar in terms of their possession of genes encoding *moaA* and *moaB*. It is possible that the presence of additional virulence related genes is associated with increased virulence of *M. tuberculosis* strains isolated during an outbreak. Translation of these genomic data requires verification in an *in vivo* infection model.

The above findings highlight the limitations of *M. tuberculosis* reference genome H37Rv in the detection of virulence-related loci. In a recent publication (O'Toole & Gautam, 2017), we reviewed the literature regarding the properties of H37Rv that possibly limits its use as a sole reference strain in tuberculosis studies.

H37Rv in itself is not a clinical isolate but a laboratory derivative of H37 which was grown from a patient's specimen at the Trudeau Institute, Saranac Lake, New York in 1905 (Kubica, Kim, & Dunbar, 1972). The phenotypic similarity of H37Rv with Robert Koch's description of TB causing bacteria (Koch, 1882) and its widespread global distribution and use led the emergence of H37Rv as a principal reference strain in diagnostics (Zhang et al., 2009) and studies on TB including vaccine design (Kaplan, 2005), therapeutics (Dutta et al., 2013), epidemiology and genomics (Mestre et al., 2011; Roetzer et al., 2013; Walker et al., 2015). The highly monomorphic feature of *M. tuberculosis* and the availability of the genome of H37Rv in the highly curated form in databases used for gene annotations, transcriptomic and proteomic studies did not warrant the search for additional reference genomes of *M. tuberculosis* until now.

However, there are reports suggesting the limitations in the robustness of using *M. tuberculosis* H37Rv as a sole reference strain in phenotypic and genotypic studies. The H37Rv strain has been associated with the failure to develop clinical TB when inoculated in low dose in the rhesus monkey (Gormus et al., 2004) and slower growth rate as compared to the isolates obtained from an infected patient (Zhang et al., 1998). De Groote and colleagues recommended the inclusion of an additional reference strain in *in vivo* drug efficacy studies (De Groote et al., 2012) while Marquina-Castillo et al concluded "the current use of H37Rv as the standard for animal models may be flawed because there were important differences in pathology caused by H37Rv" (Marquina-Castillo et al., 2009). The apparent difference in the virulence of the H37Rv strain may have occurred due to their multiple

passages in the laboratory over the longer period of time. Notable genetic differences related to the virulence of H37Rv strain include, +G insertion at -74 position upstream of the *whiB6* (Rv3862c) start codon (Solans et al., 2014), Leu152Pro substitution in *PhoR* (Rv0758) (Schreuder et al., 2015), *moa* gene containing RvD5 region (Brosch et al., 2002), and the absence of genes encoding nicotinamide adenine dinucleotide phosphate dependent oxidoreductase, and iron-regulated elongation factor (Periwal et al., 2015). Furthermore, H37Rv features the deletion of region RvD1 to RvD6 which are intact in CDC1551 (Zheng et al., 2008). It is noteworthy that CDC1151 is a highly transmissible clinical isolate of TB outbreak reported in Tennessee and Kentucky in the mid-1990s (Valway et al., 1998). A plausible conclusion regarding the difference in the virulence of H37Rv and CDC1551 may be related to their variation in gene content.

With the aim of reducing limitations associated with gene absences in the H37Rv reference genome, Okumura and colleagues developed a virtual consensus genome of *M. tuberculosis* (Okumura et al., 2015). This virtual genome was created by merging sequences of 19 *M. tuberculosis* complex strains resulting in its total length of 4,991,559 base pairs making it larger than H37Rv (NC_000962.3) by 580,027 base pairs. The possibility of inclusion of a virtual reference genome in the analysis of a clinical isolate of *M. tuberculosis* depends on its ability to detect the full repertoire of genes in the latter.

3.4 Conclusions and Recommendations

a) There are key differences between the New Zealand TB outbreak Rangipo strain and *M. tuberculosis* reference genome H37Rv in regard to virulence related genes. These were identified either as single locus variations or the presence of the additional genes in the Rangipo strain. Further studies (*in vivo* models) are required to elucidate the role of these variations to the infectivity of the Rangipo strain.

b) The Rangipo related differential carriage of virulence related genes is not an exclusive feature of this strain but has been detected in clinical isolates including *M. tuberculosis* CDC1151 strain.

c) A similarity between TB outbreak strains reported from different parts of the world was their presence of *moaA* and *moaB* genes that were absent in *M. tuberculosis* H37Rv.

d) In a preliminary analysis of whole-genome sequence data, we detected genes (with functional annotation) present in H37Rv but not in the Rangipo genome. These include, Rv1758 (cutinase 1), Rv1760 (diacylglycerol acetyltransferase), Rv2653c (toxin) and Rv2654c (antitoxin). Therefore, based on differential genome content in the outbreak strain and reference genome, we conclude that incorporation of H37Rv as a sole reference genome masks the detection of virulence related genes in *M. tuberculosis.* Based on this study, I recommend the inclusion of more than one reference genome while studying the virulence related gene in *M. tuberculosis*.

Chapter 4

Molecular Epidemiology of Tuberculosis in Tasmania, Australia

4.1 Introduction

Tuberculosis is a nationally notifiable disease across all states and territories of Australia (Toms et al., 2017). Tasmania is an island state in the southern coast of Australia with a population of approximately 0.5 million people (Australian Bureau of Statistics, 2014). The historical evidence of TB in Tasmania dates back to 1804 when Colonel David Collins documented the evidence of “consumption” in one of the settlers in Hobart (Roe, 1999). Tuberculosis remained one of the major public health problems in Tasmania during the 1940s when the state’s TB incidence rate was almost double the national rate (103.8 versus 59.3 per 100,000 population) (Communicable Disease Intelligence, 2003). By 2014, Tasmania is at the lowest end of the TB burden among states and territories in Australia with the incidence rate of 1.7/100,000 person (Toms et al., 2017). Although the published records describing the epidemiology of TB in Tasmania are limited, the disease is believed to occur as discrete cases with the majority being imported from other jurisdictions. In addition, the declaration of the eradication of bovine TB from cattle in 1975 from Tasmania as part of the Brucellosis and Tuberculosis Eradication Campaign (BTEC) in Australia has been considered to eliminate the primary source of bovine TB in humans (More, Radunz, & Glanville, 2015).

In this study, I employed whole-genome sequencing to generate the genomic data of *M. tuberculosis* complex strains isolated in Tasmania and correlated it with the public health surveillance records to report the molecular epidemiology of TB in Tasmania.

4.2 Methods

4.2.1 Study Design

Mycobacterium tuberculosis isolates reported in Tasmania were collected from the Royal Hobart Hospital located in Tasmania and Victorian Infectious Diseases Reference Laboratory, Victoria, Australia. A total of 18 cultured isolates representing 62.1% of total TB (n=29) notifications inclusive of three consecutive calendar years, 2014-2016, were studied. In addition to the clinical isolates of *M. tuberculosis*, patient data of respective cases were obtained. The samples were imported to the School of Medicine, University of Tasmania where whole-genome sequencing was performed and analysed using bioinformatic tools.

Ethical approval to conduct the study was obtained from the Tasmanian Health and Medical Human Research Ethics Committee (H0016214).

4.2.2 Culture and Drug Susceptibility Testing for *M. tuberculosis*

Samples collected from suspected cases of tuberculosis were cultured using two solid media (Brown and Buckle agar and Löwenstein-Jensen agar) and a liquid culture system using Mycobacterial Growth Indicator Tubes (MGIT) following the standard protocol (Global Laboratory Initiative,

2014). The cultured isolates were stained using the Ziehl-Neelsen staining procedure and tested for TB MPT64 antigen (a Standard Diagnostics Bioline TB MPT64 antigen test). Drug susceptibility testing of *M. tuberculosis* isolates was performed using the MGIT system (Global Laboratory Initiative, 2014).

4.2.3 Extraction of *M. tuberculosis* DNA

The genomic DNA was isolated from 1.5 mL liquid media cultures of *M. tuberculosis*. The suspension was heat treated at 80°C for 1 hour before centrifuging at 8000 rpm for 3 minutes at room temperature to obtain the cell pellet. The pelleted cells were resuspended in 200 µL phosphate-buffered saline to which 25 µL of 10 mg/mL lysozyme were added. The lysozyme treated cells were initially incubated at 37°C for 1 hour followed by incubation at 95ºC for 15 minutes. 30 µL proteinase K (10 mg/mL) was added to the mixture and then incubated for 30 minutes at 55°C. The extraction of mycobacterial DNA was then performed using a Qiagen DNeasy Blood and Tissue kit following the manufacturer's instructions. The extracted DNA was treated with 1 µL of RNase A (7000 units/mL, Qiagen) and purified using the High Pure PCR Template Preparation Kit for downstream whole-genome sequencing as discussed in Chapter 2.

4.2.4 Whole-genome Sequencing and Data Analysis

Whole-genome sequencing of the *M. tuberculosis* isolates detected in Tasmania during 2014 to 2016 was performed on an Illumina MiSeq platform as described in Chapter 2.

The data analyses pipeline involves: a) prediction of mycobacterial lineage, b) detection of drug resistance determining mutations, c) building phylogenetic relationships between the isolates and d) determining single nucleotide polymorphism difference between the isolates to infer transmission network. Mycobacterial lineages were predicted with TB Profiler database (Coll et al., 2015). The drug resistance determining mutations were detected using the PhyResSE database (Feuerriegel et al., 2015) followed by manual checking of the mapped sequences. A maximum likelihood phylogenetic inference tree rooted to *Mycobacterium canettii* was built in PhyML using the generalised time reversible (GTR) substitution model (Guindon, & Gascuel, 2003). Single nucleotide polymorphisms with respect to the reference genome H37Rv were detected as described in chapter 3, except we did not look for differential genome content and polymorphisms in genes associated with mycobacterial virulence in this chapter.

4.2.5 Identification of Tuberculosis Transmission Network in Tasmania

To establish evidence of recent TB transmission, single nucleotide polymorphic differences between the *M. tuberculosis* isolates were calculated. The isolates were associated in the same transmission network when they differed by ≤5 SNPs while >12 SNP differences excluded the isolates from the transmission network (Nikolayevskyy, Niemann & Drobniewski, 2016; Walker et al., 2013). These SNP cut-off values are likely affected by the difference in the time of sample collection, the incidence of TB in the local setting and homogeneity of *M. tuberculosis* strains in specific

geographical regions (Chad, 2014; Hatherell, 2016). Therefore, epidemiological information was analysed together with the genomic data to determine the relatedness between the cases. A possible cluster of TB was defined based on the National Tuberculosis Advisory Committee of Australia's guidelines which states that "A 'possible cluster' will be any 2 or more active cases with the same genotype of *M. tuberculosis* where temporal and geospatial association is plausible, but no direct epidemiological link is identified" (Denholm et al., 2016). *Mycobacterium tuberculosis* isolates belonging to same global lineage were analysed for SNP difference and spatiotemporal linkage. In addition, the Total Genotyping Solution for TB (TGS-TB) database (Sekizuka et al., 2015) was used to generate the *in silico* spoligotype pattern of *M. tuberculosis* isolates. This was then compared with the publicly available data of *M. tuberculosis* spoligotype reported from the patient's country of origin.

4.3 Results

4.3.1 Demographic Features of Patients Detected with TB in Tasmania

The public health surveillance records of TB cases in Tasmania during 2014 to 2016 were analysed for demographic and clinical information (Table 7) (Gautam et al., 2018). Out of a total 18 isolates of *M. tuberculosis* sequenced in this study, 7 were reported in 2016, 6 in 2015 and 5 during 2014. The major proportion (83.3%, n=15 out of 18) of the TB isolates were collected from foreign-born individuals while people born in Australia constituted 16.7% (n=3) cases. The mean age of the studied cases was

33.6 years (range, 3 months to 70 years). Pulmonary form of TB was detected in 77.7% (n=14) individuals while extrapulmonary TB (EPTB) accounted for 22.2% cases (n=4). Specimens for four individual cases of EPTB were collected as paraspinal aspirate, bone tissue, urine, and colon tissue biopsy.

Table 7. Demographic and specimen information for tuberculosis cases (*n* = 18) in Tasmania from 2014 to 2016.

Isolate	Sex	Age	Specimen	Year	Lineage	Origin
RHH2	female	23y	Sputum	2015	1	Philippines
RHH3	male	62y	Sputum	2015	3	Nepal
RHH4	female	33y	Sputum	2016	4	Thailand
RHH5	male	24y	Paraspinal aspirate	2015	1	Myanmar/Malaysia
RHH6	male	20y	Osteomyelitis	2014	3	Nepal
RHH7	female	45y	Sputum	2014	4	New Zealand
RHH8	male	20y	Sputum	2016	2	Malaysia
RHH9	male	3m	Gastric aspirate	2014	4	New Zealand
RHH10	male	23y	Sputum	2016	3	Nepal
RHH11	female	49y	Sputum	2015	3	Nepal
RHH12	male	1y	Gastric aspirate	2016	1	Philippines
RHH13	male	26y	Sputum	2015	3	Nepal
RHH14	male	1y 8m	Gastric aspirate	2015	3	Nepal
RHH15	female	67y	Sputum	2014	4	Australia
TTB1*	male	68y	Urine	2016	*M. bovis* BCG	Australia
TASMDR1	male	36y	Colon tissue	2016	2	Vietnam
TTB3	male	36y	Sputum	2016	1	Philippines
TASMB14	male	70y	Sputum	2014	*M. bovis*	Australia

*Obtained from the patient receiving BCG as a treatment for his bladder cancer

4.3.2 Lineage Distribution of *M. tuberculosis* Complex in Tasmania

The whole-genome sequence data of Tasmanian *M. tuberculosis* isolates (n=18) were genotyped using the PhyResSE and TB Profiler databases (Coll et al., 2015; Feuerriegel et al., 2015) for their distribution into different lineage types (Figure 4.1) (Gautam et al., 2018). The East-African Indian lineage 3 (*n* = 6, 33.3%) *M. tuberculosis* was predominant and was followed by the Euro-American lineage 4 (*n* = 4, 22.2%), Indo-Oceanic lineage 1 (*n* = 4, 22.2%) and East-Asian lineage 2 (n=2, 11.1%) isolates. There was 1 case of human TB caused by *M. bovis* and a case of BCG-osis in 2016.

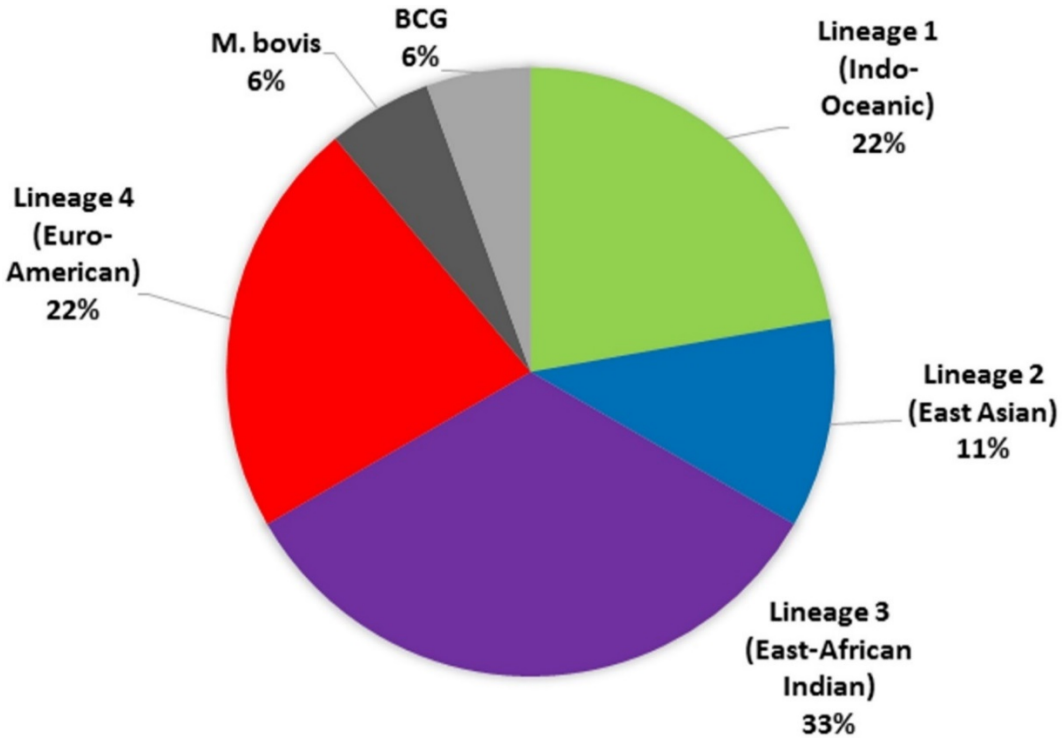


Figure 4.1. Relative frequency of *Mycobacterium tuberculosis* complex (*n* =18) lineages in Tasmania from 2014 to 2016.

Reproduced with permission from Gautam et al. PLoS ONE (2018). Copyright Gautam et al.

4.3.3 Identification of Clusters of TB in Tasmania

A phylogenetic tree of the isolates of *M. tuberculosis* collected in Tasmania was constructed using PhymL (Guindon & Gascuel, 2003). The clade specific categorisation of isolates in a maximum likelihood phylogenetic tree was in agreement with the lineage classification obtained using the PhyResSE and TB Profiler databases (Figure 4.2) (Gautam et al., 2018).

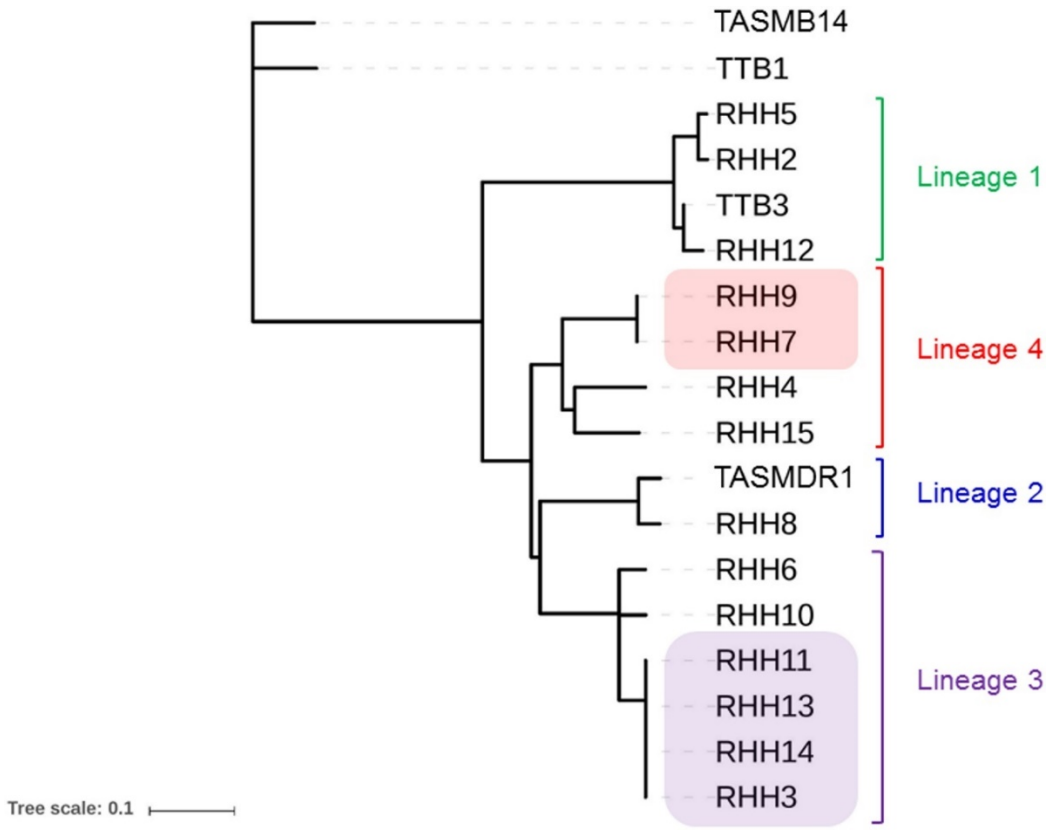


Figure 4.2. Phylogenetic relationship of *M. tuberculosis* complex isolates obtained in Tasmania from 2014 to 2016.

Generalised Time Reversible substitution model generated using PhyML (bootstraps 100). TASMB14 and TTB1 constitute isolates of *M. bovis* and *M. bovis* BCG, respectively. The Gagneux lineage numbers are indicated to

the right of each isolate on the phylogenetic tree. The colour coded boxes represent isolates representing possible clusters of TB in Tasmania.

Reproduced with permission from Gautam et al. PLoS ONE (2018). Copyright Gautam et al.

The TB transmission network in Tasmania was inferred based on the single nucleotide polymorphic differences observed between isolates of the same lineage type and phylogenetic clade. Two clusters of TB were identified in Tasmania. The first cluster (cluster 1) consisted of four isolates (RHH3, RHH11, RHH13, and RHH14) of lineage 3 *M. tuberculosis* that were separated by zero SNPs. This is evident of the TB transmission within the same network involving four individuals of cluster 1 based on the cut-off of ≤5 SNP differences between the isolates (Nikolayevskyy, Niemann, & Drobniewski, 2016; Walker et al., 2013). The patient record of four members of cluster 1 TB cases in Tasmania revealed their belonging to the same family originating from Nepal. These cases of drug susceptible pulmonary TB were detected in the gastric aspirate sample from a child and from sputum specimens collected from three adults between May and August 2015. *In silico* spoligotype data were obtained for cluster 1 isolates after importing raw sequence Fastq files into the TB profiler database (Coll et al., 2015). All four lineage 3 cluster isolates matched the CAS1_Delhi spoligotype of the Spoligotype International Type 26 (Figure 4.3) (Gautam et al., 2018). A literature search was performed to identify the reports on *M. tuberculosis* genotypes from Nepal. Malla and colleagues analysed 261 *M.*

tuberculosis isolates obtained from the cases of pulmonary TB between August 2009 to August 2010 in Nepal using spoligotyping. Lineage 3 *M. tuberculosis* were predominantly (40.6%) reported in the study with approximately 50% of these isolates belonging to CAS1_Delhi spoligotype (Malla et al., 2012).

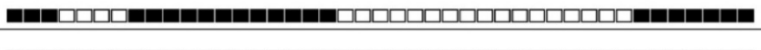
Lineage	**Spoligotype Group**	***n***	**Spoligotyping pattern** (spacers1-43)
1	EAI2	2	
	EAI2	1	
	EAI2	1	
2	Beijing	1	
	Beijing	1	
3	CAS	1	
	[a]CAS1Delhi	5	
4	T1	1	
	X1	2	
	H	1	
M.bovis	[b]X2	1	
M.bovis BCG	X	1	
Total		18	

Figure 4.3. Distribution of *in silico* generated spoligotypes across the culture-positive Tasmanian TB isolates analysed from 2014-2016.

[a]The *in silico* derived spoligotype of the four Tasmanian Lineage 3 cluster isolates (RHH3, RHH11, RHH13, RHH14) and a fifth lineage 3 isolate (RHH10) matched Spoligotype International Type 26 of the CAS1_Delhi spoligotyping family which accounted for approximately 50% of lineage 3 *M. tuberculosis* isolates in Nepal in a previous analysis (Malla B, 2012). [b]The *in silico* derived spoligotype of the *M. bovis* isolate (TASMB14) matches that of human *M. bovis* cases that were reported in other Australian

states/territories between 1977 and 1989 (Cousins, Williams, & Dawson, 1999).

In addition, the second cluster of TB (cluster 2) was detected in this study. Two isolates (RHH7 and RHH9) of lineage 4 *M. tuberculosis* obtained from cases of drug susceptible pulmonary TB during November to December 2014 exhibited zero SNP difference between one another. This transmission involved a mother and her newly born child coming from New Zealand. Although lineage 4 *M. tuberculosis* is common (70.5%, n=67/95) in New Zealand born population (Yen et al., 2013), genomic similarities were not observed between the isolates representing lineage 4 Rangipo genotype discussed in chapter 3 and members of TB cluster 2 in Tasmania.

Apart from 6 isolates of *M. tuberculosis* representing two TB clusters in Tasmania, all other isolates (n=12) were phylogenetically unrelated in terms of their SNP differences. Among these unique cases, two isolates (TTB3 and RHH12) obtained from Philippines country-of-origin individuals were closely related in terms of their SNP difference (n=74). Both the isolates belonged to lineage 1 *M. tuberculosis* which is the most common type reported in the Philippines (Gagneux, 2006; Reed, 2009). However, no epidemiological connections between these cases were apparent to categorise them as a “probable cluster” and the number of SNP differences was well above the standard 5 SNP threshold.

4.3.4 A Human Case of Bovine TB in Tasmania

An isolate (TASMB14) obtained from the TB cases detected in 2014 in Tasmania was identified as *M. bovis*. This human case of bovine TB was detected in a sputum specimen obtained from an Australia born individual aged ≥60 years with comorbidity of chronic obstructive pulmonary disease. The isolate was susceptible to first-line anti-tuberculosis drugs isoniazid and rifampicin. Whole-genome sequencing revealed the presence of pyrazinamide resistance conferring polymorphism in the Rv2043c gene (*pnc*A, Cac/Gac, H57D) and RD1 region (Rv3871 to Rv3879c) which are genomic features of *M. bovis*. Furthermore, the *in silico* spoligotype of TASMB14 matched that of earlier reported human TB cases due to *M. bovis* in Australia in 1977 and 1989 (Cousins, Williams, & Dawson, 1999) (Figure 4.3) (Gautam et al., 2018).

The genome sequence of *M. bovis* BCG isolated from the urine sample of a patient receiving BCG for the treatment of bladder cancer was mapped against M. bovis NC_008769 (BCG Pasteur1173P2). The isolate TTB1 had 99.58% template coverage with NC_008769 at the read depth of 16.97. Hence, we assumed that the BCG strain isolated from this particular patient was the same given for his treatment and did not analyse and interpret the genomic feature of TTB1 isolate further.

4.4 Discussion

In this study, I reported the first detailed analysis of the epidemiology of tuberculosis in Tasmania combining microbial whole-genome sequence and public health surveillance records. A total of 18 cultured isolates of *M.*

tuberculosis complex collected in Tasmania from 2014 to 2016, were examined. The major proportion of TB cases (83%, n=15/18) examined in this study were associated with overseas born individuals. This is comparable to the national figure of Australia where 86% of TB cases in 2014 (n=1151) were detected in foreign born patients (Toms et al., 2017). People with a history of travel or residence in high TB burden countries were the most common risk factor for TB in Australia (Toms et al., 2017).

The lineage distribution of *M. tuberculosis* complex isolates in Tasmania was reported. Four out of seven global lineages of *M. tuberculosis* were detected in Tasmania and included globally distributed Euro-American lineage 4 and East-Asian lineage 2 and relatively restricted strains belonging to Indo-oceanic lineage 1 and East-African Indian lineage 3. *M. tuberculosis* East-African Indian Lineage 3 was predominant (33.3%) in Tasmania. All of the lineage 3 cases were detected in patients originating from South-Asia where this particular lineage type is more common. A strong association between lineage of *M. tuberculosis* prevalent in patient's region of origin and isolates obtained from the migrant population in their current country of residence is well documented (Gagneux et al., 2006; Yen et al., 2013). There was no evidence supporting the spread of TB from the migrant population to the members of other communities in Tasmania. A number of studies conducted in Europe have identified that migrants are not likely to transmit TB to the local population (Barniol et al., 2009; Kamper-Jorgensen, 2012). Furthermore, a systematic review designed to study the TB transmission dynamics between the foreign-born and native-born population in European Union/ European Economic Area concluded that

“TB in a foreign-born population does not have a significant influence on TB in the native population in EU/EEA” (Sandgren et al., 2014).

Based on the whole-genome sequence data and patient variables, two possible clusters of *M. tuberculosis* were identified in Tasmania. The cases representing the largest cluster of TB in Tasmania (cluster 1) were detected within 3 months while cluster 2 cases were detected within 39 days. The *in silico* spoligotype of cluster 1 *M. tuberculosis* isolates matched the predominant spoligotype reported among the lineage 3 isolates in the patients’ country of origin, Nepal (Malla et al., 2012). The unavailability of whole-genome sequence data representing lineage 3 *M. tuberculosis* originating from Nepal limited the further analyses of transmission of cluster 1 isolates detected in Tasmania. However, based on the epidemiological and genotypic data, the origin of cluster 1 *M. tuberculosis* isolates in Tasmania could possibly be assigned to the Nepal region. In a recent publication (Gautam et al., 2018) my co-authors and I concluded that “a number of the overseas-born patients who presented with TB in Tasmania probably acquired *M. tuberculosis* infection prior to their arrival in the state or in Australia”.

Tuberculosis in migrants is more frequently detected than the native-born population in the low TB burden countries (Hanway et al., 2016), including Australia (Toms et al., 2017). The Australian Government has endorsed the World Health Organization’s post - 2015 global TB strategy directed towards ending the global epidemic of TB by reducing the current TB incidence rate by 90% and TB mortalities by 95% by 2035 (Toms et al., 2017). Tasmania should achieve a 60% drop in current TB incidence rate by 2035 to achieve

this international target. Australian TB elimination program is well supported by the availability of health care programs, robust surveillance strategies and governance frameworks (Toms et al., 2017). However, challenges persist in terms of reducing the incidence of TB in overseas born individuals. The current Australian immigration policy requires applicants aged > 11 years to undergo mandatory screening for TB if his/her intention is to stay is more than 6 months (Australian Government Department of Home Affairs, 2018). This includes a test for active TB using a chest X-ray and other diagnostic tests when indicated. An individual testing positive for active TB is barred to enter Australia until completion of treatment and declaration of being free of active TB disease (Australian Government Department of Home Affairs, 2018). It should be noted that the TB screening tests for visa purposes for Australia are not intended to detect a latent form of TB. Arguably, reactivation of latent TB is a common feature of TB among foreign born individuals in low burden high income countries (Rickset al., 2011; White, & Houben, 2014). The incidence rate of TB in Australia is predicted to decrease when the number of TB cases in the Western Pacific and South-East Asian regions decline (Toms et al., 2017). It is, therefore, the TB control policy of low burden countries (for example, Australia) should extend beyond their territories and support the high burden countries to contain the regional and global spread of TB.

In addition to cases of human infection due to *M. tuberculosis*, a case of TB was reported due to *M. bovi*s in Tasmania in 2014. This human case of bovine TB was detected in a sputum specimen obtained from an Australia born individual aged ≥60 years with a comorbidity of chronic obstructive

pulmonary disease. The isolate was susceptible to first-line anti-tuberculosis drugs isoniazid and rifampicin. The case of *M. bovis* infection was detected in an Australia born male individual approximately 40 years after the last confirmed case of bovine TB in Tasmanian in 1975 (More, Radunz, & Glanville, 2015). The *in-silico* genotype of *M. bovis* detected in Tasmania matched with the earlier reports indicating human infection due to *M. bovis* in other Australian states and territories between 1977 and 1989 (Cousins, Williams, & Dawson, 1999). No more epidemiological data were available to trace the source of pulmonary TB due to *M. bovis* in Tasmania. However, this case of *M. bovis* infection may possibly represent a reactivating case of latent TB that was acquired prior to the elimination of *M. bovis* from cattle in Tasmania. It should be noted that infection due to *M. bovis* may progress to active disease long after the exposure as explained by Cousins et al "because of the usual long incubation periods that can occur between infection and development of disease, and because of the possibility of disease reactivation, especially in elderly or immunocompromised patients, human tuberculosis caused by *M. bovis* is likely to continue to be diagnosed for many years to come" (Cousins, & Dawson, 1999). Although the human case of TB due to *M. bovis* is rare in Australia, vigilance should be maintained to control the re-emergence of this old zoonotic disease in Tasmania. Furthermore, TB due to *M. bovis* should be routinely distinguished with *M. tuberculosis* as the former is intrinsically resistant to one of the first line anti- tuberculosis drugs, pyrazinamide (Barouni et al., 2004).

4.5 Conclusions and Recommendations

a) The majority of TB cases in Tasmania is detected in overseas born individuals.

b) Although low in incidence rate, an epidemiological feature of TB in Tasmania resembles other jurisdictions in terms of the presence of clustered cases.

c) The TB control strategy in Tasmania should be more focused on detecting cases in individuals originating from high TB burden countries.

d) The detection of a human case of bovine TB in Tasmania cautions towards maintaining vigilance to control re-emergence of bovine TB in cattle and human.

Chapter 5

Tracing the Source of the First- Case of Multi-Drug Resistant Tuberculosis in Tasmania

5.1 Drug Resistant Tuberculosis

Multi-drug resistant (MDR) tuberculosis refers to TB caused by an isolate of *M. tuberculosis* that is resistant to at least the two first line anti-TB drugs, isoniazid and rifampicin (World Health Organization, 2018i). The World Health Organization recommends MDR-TB treatment regime for rifampicin mono-resistant (RR) isolates (World Health Organization, 2016a) as RR is considered as a marker of MDR-TB in more than 90% cases (Drobniewski & Pozniak, 1996). An estimated 558,000 cases globally were resistant to rifampicin (RR-TB) in 2017 (World Health Organization, 2018d). Eighty-two percent (n=457,560) of these RR-TB cases were identified as MDR-TB. India (24%), China (13%) and the Russian Federation (10%) collectively shared approximately half (47%) of global MDR-TB cases in 2017 (World Health Organization, 2018d). Of the total MDR-TB cases, 8.5% (95% CI, 6.2-11%) were extensively drug resistant TB (World Health Organization, 2018d). In terms of treatment, drug susceptible TB cases were more successfully cured as compared to MDR and XDR form of TB in 2016 (82% versus 54% and 30% respectively) (World Health Organization, 2018d).

In Australia, an estimated 25 (14-36) cases of MDR/RR-TB were notified among pulmonary TB cases in 2017 (World Health Organization, 2018). A total of 23 cases were bacteriologically confirmed for MDR/RR-TB, of which,

23 patients were enrolled for treatment (World Health Organization, 2018). In terms of treatment success, 78% of the new and relapse case, 75% of MDR/RR-TB and 100% XDR-TB was successfully achieved in their size of 1371, 32 and 2 cohorts respectively (World Health Organization, 2018). In terms of patients' country of origin, approximately 88% (n=15 out of total 17 cases) of MDR/XDR-TB patients in 2014 in Australia were foreign born individuals (Toms et al., 2017).

Tasmania exhibited the lowest TB incidence rate (1.7 per 100,000 population) among 8 Australian states or territories in 2014 (Toms et al., 2017). Here, I utilized whole-genome sequencing to define the genomic features and track the transmission of the first case of MDR-TB sequenced during the analysis of Tasmanian *M. tuberculosis* isolates as described in chapter 4 (section 4.2.4).

5.2 Methods

5.2.1 Whole-genome Sequencing of the First Isolate of MDR-TB in Tasmania

One of the Tasmanian-collected TB isolates from 2014-2016 exhibited phenotypic drug resistance to isoniazid, rifampicin, ethambutol and pyrazinamide as determined by Victoria Infectious Diseases Reference Laboratory, Melbourne Australia (VIDRL). Whole-genome sequencing of this isolate was performed on an Illumina MiSeq platform as described in Chapter 2.

The raw sequence data were analysed for lineage type using the TB profiler database (Coll et al., 2015) and mutations associated with drug resistance were detected in PhyResSE database (Feuerriegel et al., 2015). This was followed by manual checking of the sequence for variants associated with drug resistance in *M. tuberculosis*. In addition, the Fastq files were imported onto the Geneious software suite (R 9.5) (Kearse et al., 2012) and analysed as described in Chapter 3.

Ethical approval to conduct the study was obtained from the Tasmanian Health and Medical Human Research Ethics Committee (H0016214).

5.2.2 Consensus Genome Assembly of First MDR-TB Isolate in Tasmania

The paired end raw Fastq reads of the MDR-TB isolate reported for the first time in Tasmania were generated using an Illumina MiSeq sequencer. The Burrows-Wheeler Aligner (Li & Durbin, 2010) was used to map the Fastq files with the *M. tuberculosis* reference genome H37Rv (NC_000962). After mapping, the contigs were de novo assembled using the SPAdes assembler (version 3.7) (Bankevich et al., 2012) and ordered according to *M. tuberculosis* H37Rv using ABACAS (Assefa et al., 2009) for their deposition in public databases. The number of variants in the MDR-TB strain with respect to the H37Rv reference genome were determined using the SAM-tools analyses suite. These variants were then annotated using the SnpEff analyses platform (Cingolani et al., 2012). To identify the presence of drug resistance determining mutations, paired end raw Fastq files were imported

onto publicly available PhyResSE database (Feuerriegel et al., 2015). The whole-genome shotgun project of the first MDR-TB isolate detected in Tasmania was submitted for deposition at DNA Data Bank of Japan, European Molecular Biology Laboratory and National Institute of Health genetic sequence database, GenBank and is available under accession number NTFG00000000.

5.2.3 Tracing the Source of the First MDR-TB Case in Tasmania

The clinical and epidemiological information of the first case of MDR-TB detected in Tasmania were obtained from the treating physician at Launceston General Hospital, Tasmania. To establish a transmission network, a cut off of 5 or less single nucleotide polymorphisms between the isolates was used as described previously (Gautam et al., 2018; Walker et al., 2013). The difference in time of sample collection from different sources, TB incidence rate in the local setting and homogeneity of a particular *M. tuberculosis* strain in a specific geographical location may influence the SNP distance calculation to determine the transmission chain (Hatherell et al., 2016; Migliori et al., 2008). To minimize potential bias that may occur when genomic data are solely used to infer recent transmission (Hatherell et al., 2016 and Walker et al., 2013), we analysed the whole-genome sequence data together with epidemiological information to confirm recent transmission of MDR-TB in Tasmania.

5.3 Results

5.3.1 Case of Drug Resistant TB in Tasmania

Tasmania had been free of the multi-drug resistant form of TB until the first case of MDR-TB was diagnosed in 2016 in a Vietnamese-born individual at Launceston General Hospital. The patient was a 37-year-old male student who tested positive for latent TB in an interferon gamma release assay in February 2016 but exhibited no symptoms of pulmonary TB and reported negative in chest X-ray investigation and sputum smear microscopy and culture. A few months later, the patient presented with a complaint of abdominal pain consistent with colitis. A colon tissue biopsy sample was obtained and examined for microbiological features of *M. tuberculosis* infection. The sample was negative for smear microscopy but the growth of tubercle bacilli in culture evidenced the diagnosis of extra-pulmonary TB. Drug susceptibility profiling of the isolate was performed at VIDRL using Mycobacteria Growth Indicator Tube (MGIT) system. The isolate was recorded as resistant to isoniazid, rifampicin, ethambutol and pyrazinamide and susceptible to ethionamide, amikacin, capreomycin, kanamycin, ofloxacin and moxifloxacin by VIDRL. This confirmed the first reported case of multi-drug resistant isolate of *M. tuberculosis* in Tasmania (TASMDR1).

5.3.2 Genomic Features of TASMDR1

The report on the genomic feature of TASMDR1 was published (Gautam, Mac Aogáin, & O'Toole, 2017) in Genome Announcements™, a publication

of American Society for Microbiology. This particular analysis was done to generate data on TASMDR1 and deposit it in the public databases. In the publication (Gautam, Mac Aogáin, & O'Toole, 2017), we mentioned "A total of 2,860,297 paired-end reads were mapped to the publicly available annotated genome of *M. tuberculosis* reference strain H37Rv (GenBank accession number NC_000962.3) (Cole et al., 1998) by Burrows-Wheeler alignment (Li & Durbin, 2010). This yielded an average read depth of 65.5-fold, covering 97.8% of the H37Rv genome. A 4,230,496-bp draft genome assembly of 220 contigs was assembled *de novo* using the SPAdes assembler (v3.7) (Bankevich et al., 2012). A total of 1,553 variant sites were identified relative to the H37Rv genome and consisted of 1,408 single-nucleotide variants (SNVs) and 145 insertions/deletions (indels size range, 1 to 501 base pairs). Of the variants, 881 were nonsynonymous; of these, 784 were SNVs and 97 were insertions/deletions. The genome of TASMDR1 displayed high-confidence single-nucleotide polymorphisms in genes correlating with antimicrobial drug resistance when analysed using the PhyResSE database (Feuerriegel et al., 2015). These included high-confidence mutations in the *katG* gene (aGc/aCc, S315T) and *rpoB* gene (gAc/gGc, D435G; tCg/tTg, S450L), which underlie *M. tuberculosis* resistance to isoniazid and rifampicin, respectively (Eldholm et al., 2015; Rodwell et al., 2014). These data establish the genetic bases of the MDR phenotype exhibited by strain TASMDR1. Additional mutations were detected in the *embB* gene (Atg/Gtg, M306V) and *pncA* gene (cCg/cTg, P62L) that are associated with resistance to ethambutol and pyrazinamide, respectively (Cuevas-Córdoba et al., 2015; Miotto et al., 2014; Simons et

al., 2014; Starks et al., 2009)". In addition to the drug resistance determining mutations reported by a clinical laboratory, this study identified genomic evidence of resistance of TASMDR1 towards streptomycin (*rrs* (MTB000019)) (Table 8) (Gautam et al., 2018). The PhyResSE and TB Profiler databases (Coll et al., 2015; Feuerriegel et al., 2015) predicted TASMDR1 as an isolate of *M. tuberculosis* belonging to the Beijing sub-lineage of *M. tuberculosis* East Asian Lineage 2.

This whole-genome shotgun project is available at DDBJ/ENA/GenBank under the accession number NTFG00000000. The version that has been described in the publication is NTFG01000000 (Gautam, Mac Aogáin, & O'Toole, 2017).

Table 8. Drug resistance determining mutations present in TASMDR1.

Isolates	Drug	Gene	Locus tag	Mutation	Genome location	Substitution
TASMDR1	Rifampicin	*rpoB*	Rv0667	gAc/gGc, tCg/tTg	761110, 761155	D435G, S450L
	Isoniazid	*katG*	Rv1908c	aGc/aCc	2155168	S315T
	Pyrazinamide	*pncA*	Rv2043c	cCg/cTg	2289057	P62L
	Ethambutol	*embB*	Rv3795	Atg/Gtg	4247429	M306V
	Streptomycin	*rrs*	MTB000019	a/c	1472359	a514c

5.3.3 Tracing the Source of TASMDR1

The treating physician at Launceston General Hospital became aware that the Tasmanian MDR-TB patient was in close contact with his family member detected with multi-drug resistant pulmonary TB in Vietnam in 2012 prior to this arrival in Australia. I travelled to Cho Ray Hospital, Ho Chi Minh city, Vietnam to study the contact case in detail. The culture-based laboratory diagnosis (MGIT system) of an isolate obtained from the contact case was recorded as resistant to isoniazid, rifampicin, ethambutol, pyrazinamide, and streptomycin. In addition, genomic DNA of the *M. tuberculosis* isolate obtained from the contact case was shipped to the University of Tasmania for further analyses. This isolate obtained from Vietnam was labelled as VTB1.

Whole-genome sequencing was performed on genomic DNA of VTB1 and data were analysed as described in chapter 2. The genomic data was interpreted together with the available epidemiological and clinical information (Table 9). The drug resistance profile and lineage typing of VTB1 were identical to TASMDR1, i.e. VTB1 belonged to East Asian L2 (Beijing) and conferred exactly the same drug resistance determining mutations as of TASMDR1 i.e. isoniazid, rifampicin, pyrazinamide, ethambutol and streptomycin. In addition, no drug resistance determining mutations were detected for ethionamide, amikacin, capreomycin, kanamycin, ofloxacin and moxifloxacin. To build a transmission network, single nucleotide polymorphic differences between TASMDR1 and VTB1 were calculated. The isolates were detected to separate from each other by zero SNP differences which is strongly indicative of their belonging to the

same transmission network. With an established epidemiological link of household contact and diagnosis of VTB1 case as an infectious form (i.e. pulmonary TB) of TB prior to TASMDR1, my co-authors and I in our recent publication (Gautam et al., 2019) paper concluded that “the Tasmanian patient contracted MDR-TB strain from his household contact early before his arrival to Tasmania which then reactivated from the latency and clinically presented as a case of multi-drug resistant extra-pulmonary TB in 2016 in Tasmania”.

Table 9. Epidemiological and Clinical Profile of the First Case of MDR-TB Detected in Tasmania (TASMDR1) and its Contact (VTB1).

Demographic and clinical details	**TASMDR1**	**VTB1**
Sex	Male	Female
Age*	37	31
Date of Detection	2016	2012
Country of Origin	Vietnam	Vietnam
Location of Diagnosis	Launceston, Tasmania, Australia	Ho Chi Minh City, Vietnam
Epidemiological link	Household family members	
Clinical case	Extra-pulmonary TB	Pulmonary TB
Isolate	*M. tuberculosis*	*M. tuberculosis*
Lineage	East Asian L2 (Beijing)	East Asian L2 (Beijing)
Drugs Resistant	Isoniazid, rifampicin, pyrazinamide, ethambutol and streptomycin	Isoniazid, rifampicin, pyrazinamide, ethambutol and streptomycin
Number of SNP Difference	Zero with respect to each other	

*Based on the year of diagnosis

5.4 Discussion

In this study, I analysed the whole-genome sequence data and patient information to genetically and epidemiologically link the first case of MDR-TB in Tasmania to its household contact in Vietnam. Australia is a low TB burden country with MDR-TB occurring in less than 2% of incident cases (Toms et al., 2017). The majority of MDR/XDR-TB cases in Australia are diagnosed among foreign born individuals, for example, 15 out of 17 cases of MDR-TB were reported in overseas born population in 2014 (Toms et al., 2017). Francis and colleagues retrospectively reviewed the laboratory confirmed cases of MDR-TB in Australia reported during 1998-2012 and found that the majority of cases occurred in overseas-born individuals as compared to people born in Australia (94.7% versus 5.3% respectively, n=244) (Francis et al., 2018).

In this study, I reported the genomic feature of the first case of MDR-TB detected in Tasmania (TASMDR1) in 2016. The source tracing of TASMDR1 was done using the epidemiological and whole-genome sequence data. Lalor and colleagues recently recommended the combined use of WGS and epidemiological information to investigate the clusters of MDR-TB (Lalor et al., 2018). The TASMDR1 patient originated from Vietnam and the source of this case was traced to his household contact in Vietnam. Vietnam ranked second (n=140) to India (n=223) in terms of TB patient's country of origin in Australia (Toms et al., 2017). Vietnam is one of the high TB and TB/HIV burden countries (World Health Organization, 2018h) where 4.1% of new and 17% of previously treated cases were detected as MDR/RR-TB. It is unlikely that the MDR-TB isolate in our study

was imported in the form of an active MDR-TB case from Vietnam to Tasmania. This is because the student detected with TASMDR1 would have undergone the mandatory pre-arrival health check-up including screening for active pulmonary TB during the process of his visa application to Australia. Furthermore, the patient did not exhibit clinical signs or symptoms of active pulmonary TB in early 2016 post arrival in Tasmania.

Chest X-ray (CXR) imaging is a primary tool to screen for active TB in people entering Australia. The CXR is unreliable in detecting *M. tuberculosis* infection in extrapulmonary sites away from the lung periphery. Pollett and colleagues reported a median time of 5 years (0-41 years) to detect cases of extrapulmonary TB in people migrating from high burden countries to Australia and concluded that “Vigilance for extrapulmonary TB needs to remain high in those moving from high prevalence countries to Australia, even decades after immigration”. The authors further highlighted the challenge in the diagnosis of extra pulmonary TB in their well-equipped setting of Centre for Infectious Diseases and Microbiology, Western Sydney Local Health District where half of the extra-pulmonary TB did not produce microbiological evidence of the disease (Pollett et al., 2016). Globally, 14% of all TB cases are detected in extrapulmonary sites (World Health Organization, 2018d) and the complexity in the clinical and laboratory diagnosis of these cases pose a challenge for TB control programs, in particular, their spread from high to low TB burden settings in latent form.

Clinically, the majority of TB cases among the foreign-born population in industrialised countries represent reactivation of LTBI rather than the

continuation of an existing case of active TB (Ricks et al., 2011; White & Houben, 2014). Therefore, detection of both the active and/or latent TB infection and their early management is the major focus in controlling TB in migrants and its spread. In addition to the diagnostic limitation, inability to clinically identify if *M. tuberculosis* associated with latent TB is susceptible or resistant to drugs adds uncertainties in selecting the effective treatment regime (Norbis et al., 2014). Poor compliance towards TB treatment in refugees and immigrants has been related to their issues of privacy, distress, shame and possible isolation (Wieland et al., 2012) and perceived side effects of medication (Wieland et al., 2012). A systematic review of a national estimate of MDR-TB risk factors in 6 European nations associated overseas birth as a risk of MDR-TB acquisition (odds ratio (OR) 2.46; 95%CI, 1.86-3.24). As explained in chapter 4 “It is, therefore, the TB control policy of low burden countries should extend beyond their territories and support the high burden countries to contain the regional and global spread of TB”.

5.5 Conclusion and Recommendations

In a recent publication highlighting the intercontinental transmission of MDR-TB in Australia (Gautam et al., 2019), my co-authors and I concluded that “despite the immense prevalence of LTBI, there are few reports in the literature describing the translocation of the multi-drug resistant form of tuberculosis from one jurisdiction to another as a latent infection, and its subsequent emergence as active MDR-TB in a new host country. This type of international movement of TB is difficult to detect with existing pre-

immigration TB screening practices that are reliant upon a chest x-ray-based diagnosis of a current or previous episode of pulmonary TB. The silent international movement of MDR-TB in latent form, as is suspected in this case, is an area of concern". Hence, caution needs to be maintained through tracing the source of MDR-TB cases and implementing TB control strategies to facilitate the early detection of drug resistance TB and limit their spread particularly in low prevalence setting like Tasmania.

Chapter 6

Summary

Tuberculosis remains a global public health emergency. The management of TB is largely impeded due to the limited protective efficacy of the BCG vaccine, diagnostic delay, and treatment failure either due to patient non-compliance or poor allocation of resources. However, on a positive note, the global incidence of TB is decreasing by 2% per year and the introduction of new technologies have aided in the diagnosis and epidemiological typing of disease. My thesis is based on the application of whole-genome sequencing in functional and epidemiological typing of tuberculosis. Initially, I described an in-house protocol, adapted at Trinity College Dublin, to obtain whole-genome sequence data of *M. tuberculosis*. Here, we modified the manufacturer's protocol in terms of consumables and library normalization step. A major advantage of doing so was the reduction in cost associated with consumables while obtaining quality data for downstream analyses. However, we did not validate the proposed method against any published protocol which is one of the limitations of the study. For the first section of my thesis, I hypothesised that virulence related genes of *M. tuberculosis* complex isolate capable of causing outbreaks vary at a certain level with respect to the reference strain H37Rv. To test my hypothesis, I performed functional genomic analyses of the New Zealand TB outbreak "Rangipo" strain. In this study, I reported differential carriage of virulence related genes and amino-acid changing mutations in genes previously related to

mycobacterial virulence. This further led to conclude that mapping whole-genome sequence data with a widely adopted reference strain H37RV alone masks the detection of virulence related genes. Therefore, I recommended the inclusion of additional reference strains (for example CDC1551) while analysing the whole-genome sequence data of *M. tuberculosis*.

In the second part, I performed molecular epidemiological typing of *M. tuberculosis* isolated in Tasmania during three consecutive calendar years (2014-2016). Here, I described for the first time a comprehensive report detailing the molecular epidemiology of TB in Tasmania. I reported the predominance of TB among non-Australia born individuals, lineage 3 as a dominant type and the detection of two clusters of TB cases in Tasmania. Furthermore, a human case of infection due to *M. bovis* was described in the study. For the third aim, I implemented WGS to trace the source of the first case of MDR-TB reported in Tasmania. The origin of TASMDR1 was located to Vietnam and evidence for its transmission to Tasmania in the latent stage was discussed. While addressing aim 2 and aim 3, I concluded that the epidemiological feature of TB in low burden setting of Tasmania is no different than other jurisdictions. Hence, vigilance should be maintained to timely detect the cases, in particular, migrant population coming from high burden countries.

To summarize, this thesis demonstrated the application of whole-genome sequencing to study the functional genomics and epidemiology of *M. tuberculosis*.

References

Abdallah, AM, Hill-Cawthorne, GA, Otto, TD, Coll, F, Guerra-Assunção, JA, Gao, G, Naeem, R, Ansari, H, Malas, TB, Adroub, SA, Verboom, T, Ummels, R, Zhang, H, Panigrahi, AK, McNerney, R, Brosch, R, Clark, TG, Behr, MA, Bitter, W & Pain, A 2015, 'Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations', *Sci Rep*, vol.5, doi: 10.1038/srep15443.

Abraham, EP, Chain, E, Fletcher, CM, Gardner, AD, Heatley, NG, Jennings, MA, & Florey, HW 1941, 'Further observations on penicillin', *Lancet*, vol. 238, no. 6155, pp. 177-189.

Afgan, E, Sloggett, C, Goonasekera, N, Makunin, I, Benson, D, Crowe, M, Gladman, S, Kowsar, Y, Pheasant, M, Horst, R, & Lonie, A 2015, 'Genomics virtual laboratory: a practical bioinformatics workbench for the cloud', *PLoS One*, vol. 10, no. 10, doi: 10.1371/journal.pone.0140829.

Alangaden, GJ, Kreiswirth, BN, Aouad, A, Khetarpal, M, Igno, FR, Moghazeh, SL, Manavathu, EK, Lerner, SA 1998, 'Mechanism of resistance to amikacin and kanamycin in *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 42, no. 5, pp. 1295-1297.

Alkabab, YM., Enani, MA., Indarkiri, NY, & Heysell, SK 2018, 'Performance of computed tomography versus chest radiography in patients with pulmonary tuberculosis with and without diabetes at a tertiary hospital in Riyadh, Saudi Arabia', *Infect Drug Resist*, vol.11, pp. 37-43.

Allix, C, Supply, P & Fauville-Dufaux, M 2004, 'Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis', *Clin Infect Dis*, vol. 39, no. 6, pp. 783-789.

Allix-Beguec, C, Wahl, C, Hanekom, M, Nikolayevskyy, V, Drobniewski, F, Maeda, S, Campos-Herrero, I, Mokrousov, I, Niemann, SK, Kontsevaya, I, Rastogi, N, Samper, S, Sng, LH, Warren, RM, & Supply P 2014, 'Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci for subtyping of *Mycobacterium tuberculosis* Beijing isolates', *J Clin Microbiol*, vol. 52, no.1, pp.164-172.

Altschul, SF, Gish, W, Miller, W, Myers, EW & Lipman, DJ 1990, 'Basic local alignment search tool', *J Mol Biol*, vol. 215, no. 3, pp. 403-410.

Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W & Lipman, DJ, 1997, 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Res*, vol. 25, no. 17, pp. 3389-3402.

Amlerova, J, Bitar, I, & Hrabak, J 2018', 'Genotyping of *Mycobacterium tuberculosis* using whole-genome sequencing', *Folia Microbiologica*, vol. 63, no. 5, pp. 537-545.

Andrews S, 2010, 'FastQC: a quality control tool for high throughput sequence data'. [Electronic edition] Retrieved from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arbelaez, MP, Nelson, KE, & Munoz, A 2000, 'BCG vaccine effectiveness in preventing tuberculosis and its interaction with human immunodeficiency virus infection', *Int J Epidemiol*, vol. 29, no. 6, pp.1085-1091.

Arnold, C 2007, 'Molecular evolution of *Mycobacterium tuberculosis*', Clin *Microbiol Infect*, vol. 13 no. 2, pp. 120-128.

Arnold, C, Thorne, N, Underwood, A, Baster, K, & Gharbia, S 2006, 'Evolution of short sequence repeats in *Mycobacterium tuberculosis*', *FEMS Microbiol Lett*, vol. 256, no. 2, pp. 340-346.

Assefa, S, Keane, TM, Otto, TD, Newbold, C & Berriman, M 2009, 'ABACAS: algorithm-based automatic contiguation of assembled sequences', *Bioinformatics*, vol. 25, no. 15, pp. 1968-1969.

Atkinson J, Chartier, Y, Pessoa-Silva, CL, Jensen, P, Li, Y & Seto WH 2009, 'Natural Ventilation for Infection Control in Health-Care Settings', [Electronic edition] Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK143281/

Australian Bureau of Statistics 2014, 'Population by Age and Sex, Regions of Australia', [Electronic edition] Retrieved from: http://www.abs.gov.au/ausstats/abs@.nsf/Products/3235.0~2011~Main+Features~Tasmania?OpenDocument

Australian Government Department of Home Affairs 2018, ' What health examinations you need', [Electronic edition] Retrieved from: https://immi.homeaffairs.gov.au/help-support/meeting-our-requirements/health/what-health-examinations-you-need

Azhikina, T, Gvozdevsky, N, Botvinnik, A, Fushan, A, Shemyakin, I, Stepanshina, V, Lipin M, Barry, C3rd & Sverdlov, E 2006, 'A genome-wide sequence-independent comparative analysis of insertion–deletion polymorphisms in multiple *Mycobacterium tuberculosis* strains', *Res in Microbiol*, vol. 157, no. 3, pp. 282-290.

Bainomugisa, A, Lavu, E, Hiashiri, S, Majumdar, S, Honjepari, A, Moke, R, Dakulala, P, Hill-Cawthorne, GA, Pandey, S, Marais, BJ, Coulter, C, & Coin, L 2018, 'Multi-clonal evolution of multi-drug-resistant/extensively drug-resistant *Mycobacterium tuberculosis* in a high-prevalence setting of Papua New Guinea for over three decades', *Microb Genom*, vol. 4, no. 2, doi: 10.1099/mgen.0.000147.

Banerjee, A, Dubnau, E, Quemard, A, Balasubramanian, V, Um, KS, Wilson, T, Collins, D, de Lisle, G & Jacobs, WR, 1994, 'inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*', *Science*, vol. 263, no. 5144, pp. 227-230.

Bankevich, A, Nurk, S, Antipov, D, Gurevich, AA, Dvorkin, M, Kulikov, AS, Lesin, VM, Nikolenko, SI, Pham, S, Prjibelski, AD, Pyshkin, AV, Sirotkin, AV, Vyahhi, N, Tesler, G, Alekseyev, MA, Pevzner, PA 2012, 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *J Comput Biol*, vol. 19, no. 5, pp. 455-477.

Barberis, I, Bragazzi, NL, Galluzzo, L & Martini, M 2017, 'The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus', *J Prev Med Hyg*, vol. 58, no. 1, E9-E12.

Barbier, M & Wirth, T 2016, 'The Evolutionary History, Demography, and Spread of the *Mycobacterium tuberculosis* Complex', *Microbiol Spectr*, vol. 4, no.4, doi: 10.1128/microbiolspec.TBTB2-0008-2016.

Barnes, PF & Cave, MD 2003, 'Molecular epidemiology of tuberculosis', *N Engl J Med*, vol. 349, no. 12, pp.1149-1156.

Barniol, J, Niemann, S, Louis, VR, Brodhun, B, Dreweck, C, Richter, E, Becher, H, Haas, W, Junghanss, T 2009, 'Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant sub-populations', *BMC Infect Dis*, vol. 9, no. 197, doi: 10.1186/1471-2334-9-197.

Barouni, AS, Augusto, CJ, Lopes, MTP, Zanini, MS & Salas CE 2004, 'A pncA polymorphism to differentiate between Mycobacterium bovis and *Mycobacterium tuberculosis*', *Mol Cell Probes*, vol. 18, no. 3, pp. 167-170.

Barry, CE 3rd, Boshoff, HI, Dartois, V, Dick, T, Ehrt, S, Flynn, J, Schnappinger, D, Wilkinson, RJ, Young, D 2009, 'The spectrum of latent tuberculosis: rethinking the biology and intervention strategies', *Nat Rev Microbiol*, vol. 7, no. 12, pp. 845-855.

Beckert, P, Hillemann, D, Kohl, T, Kalinowski, J, Richter, E, Niemann, S & Feuerriegel, S 2012, 'rplC T460C identified as a dominant mutation in linezolid-resistant Mycobacterium 630 tuberculosis strains', *Antimicrob Agents Chemother*, vol. 56, no.5, pp. 2743-2745.

Bhat, RM, & Prakash, C 2012, 'Leprosy: An Overview of Pathophysiology. Interdisciplinary Perspectives on Infectious Diseases', *Interdiscip Perspect Infect Dis*, vol. 2012, doi:10.1155/2012/181089.

Bjorn-Mortensen, K, Soborg, B, Koch, A, Ladefoged, K, Merker, M, Lillebaek, T, Andersen AB, Niemann, S & Kohl, TA 2016, 'Tracing *Mycobacterium tuberculosis* transmission by whole-genome sequencing in a high incidence setting: a retrospective population-based study in East Greenland', *Sci Rep*, vol. 6, doi:10.1038/srep33180.

Black, GF, Weir, RE, Floyd, S, Bliss, L, Warndorff, DK, Crampin, AC, Ngwira, B, Sichali, L, Nazareth, B, Blackwell, JM, Branson, K, Chaguluka, SD, Donovan, L, Jarman, E, King, E, Fine, PE & Dockrell, H M 2002, 'BCG-induced increase in interferon-gamma response to mycobacterial antigens and efficacy of BCG vaccination in Malawi and the UK: two randomised controlled studies', *Lancet*, vol. 359, no. 9315, pp. 1393-1401.

Bloemberg, GV, Gagneux, S, & Böttger, EC 2015, 'Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis', *N Engl J Med*, vol. 373, no. 20, pp. 1986-1988.

Bradley, P, Gordon, NC, Walker, TM, Dunn, L, Heys, S, Huang, B, Earle, S, Pankhurst, LJ, Anson, L, de Cesare, M, Piazza, P, Votintseva, AA, Golubchik, T, Wilson, DJ, Wyllie, DH, Diel, R, Niemann, S, Feuerriegel S, Kohl T, Ismail, N, Omar, SV, Smith, EG, Buck, D, McVean, G, Walker, AS, Peto, TE, Crook, DW & Iqbal, Z 2015, 'Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and *Mycobacterium tuberculosis*', *Nat Commun*, vol. 6, doi:10.1038/ncomms10063.

Brandt, L, Feino Cunha, J, Weinreich Olsen, A, Chilima, B, Hirsch, P, Appelberg, R, & Andersen, P 2002, 'Failure of the Mycobacterium bovis

BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis', *Infect Immun*, vol. 70, no. 2, pp. 672-678.

Brosch, R, Gordon, SV, Marmiesse, M, Brodin, P, Buchrieser, C, Eiglmeier, K, Garnier, T, Gutierrez, C, Hewinson, G, Kremer, K, Parsons, LM, Pym, AS, Samper, S, van Soolingen, D & Cole, ST 2002, 'A new evolutionary scenario for the *Mycobacterium tuberculosis* complex', *Proc Natl Acad Sci USA*, vol. 99, no. 6, pp. 3684-3689.

Brown, AC, Bryant, JM, Einer-Jensen, K, Holdstock, J, Houniet, DT, Chan, JZ, Depledge, DP, Nikolayevskyy, V, Broda, A, Stone, MJ, Christiansen, MT, Williams, R, McAndrew, MB, Tutill, H, Brown, J, Melzer, M, Rosmarin, C, McHugh, TD, Shorten, RJ, Drobniewski, F, Speight, G, & Breuer, J 2015, 'Rapid Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples', *J Clin Microbiol*, vol. 53, no. 7, pp. 2230-2237.

Bryant, JM, Schürch, AC, van Deutekom, H, Harris, SR, de Beer, JL, de Jager, V, Kremer, K, van Hijum, SA, Siezen, RJ, Borgdorff, M, Bentley, SD, Parkhill, J & van Soolingen, D 2013, 'Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole-genome sequencing data', *BMC Infect Dis*, vol. 13, no. 1, doi:10.1186/1471-2334-13-110.

Buermans, HP & den Dunnen, JT 2014, 'Next generation sequencing technology: Advances and applications', *Biochim Biophys Acta*, vol. 1842, no. 10, pp.1932-1941.

Cannas, A, Mazzarelli, A, Di Caro, A, Delogu, G & Girardi, E 2016, 'Molecular Typing of *Mycobacterium tuberculosis* Strains: A Fundamental Tool for Tuberculosis Control and Elimination', *Infect Dis Rep*, vol. 8, no. 2, doi:10.4081/idr.2016.6567

Cardoso Oelemann, M, Gomes, HM, Willery, E, Possuelo, L, Batista Lima, KV, Allix-Béguec, C, Locht, C, Goguet de la Salmonière, YO, Gutierrez, MC, Suffys, P & Supply, P 2011, 'The Forest behind the Tree: Phylogenetic Exploration of a Dominant *Mycobacterium tuberculosis* Strain Lineage from a High Tuberculosis Burden Country', *PLoS One*, vol. 6, no. 3, doi: 10.1371/journal.pone.0018256.

Casali, N, Nikolayevskyy, V, Balabanova, Y, Harris, SR, Ignatyeva, O, Kontsevaya, I, Corander, J, Bryant, J, Parkhill, J, Nejentsev, S, Horstmann, RD, Brown, T & Drobniewski, F 2014, 'Evolution and transmission of drug-resistant tuberculosis in a Russian population', *Nat Genet,* vol. 46, no. 3, pp.279-86.

Centers for Disease Control and Prevention 2012, 'Guidelines for Screening for Tuberculosis Infection and Disease during the Domestic Medical Examination for Newly Arrived Refugees', [Electronic edition] Retrieved from: https://www.cdc.gov/immigrantrefugeehealth/pdf/domestic-tuberculosis-refugee-health.pdf

Centers for Disease Control and Prevention 2018, 'Transmission of Measles', [Electronic edition] Retrieved from: https://www.cdc.gov/measles/about/transmission.html

Chaisson, RE 2003, ‘Tuberculosis Chemotherapy: still a double-edged sword’, *Am J Respir Crit Care Med*, vol. 167, no. 11, pp. 1461-1462.

Chaisson, RE, Schecter, GF, Theuer, CP, Rutherford, GW, Echenberg, DF & Hopewell, PC 1987, ‘Tuberculosis in patients with the acquired immunodeficiency syndrome. Clinical features, response to therapy, and survival’, *Am Rev Respir Dis*, vol. 136, no. 3, pp. 570-574.

Cingolani, P, Platts, A, Wang, LL, Coon, M, Nguyen, T, Wang, L, Land, SJ, Lu, X & Ruden, DM 2012, ‘A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3’, *Fly*, vol. 6, no. 2, pp. 80-92

Click, ES, Moonan, PK, Winston, CA, Cowan, LS, & Oeltmann, JE 2012, ‘Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis’, *Clin Infect Dis*, vol. 54, no. 2, pp. 211-219.

Colangeli, R, Arcus, VL, Cursons, RT, Ruthe, A, Karalus, N, Coley, K, Manning, SD, Kim, S, Marchiano, E & Alland, D 2014, ‘Whole-genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans’, *PLoS One*, vol. 9, no.3, doi: 10.1371/journal.pone.0091024.

Cole, S, Brosch, R, Parkhill, J, Garnier, T, Churcher, C, Harris, D, Gordon, SV, Eiglmeier, K, Gas, S, Barry, CE 3rd, Tekaia, F, Badcock, K, Basham, D, Brown, D, Chillingworth, T, Connor, R, Davies, R, Devlin, K, Feltwell, T, Gentles, S, Hamlin, N, Holroyd, S, Hornsby, T, Jagels, K, Krogh, A, McLean, J, Moule, S, Murphy, L, Oliver, K, Osborne, J, Quail, MA, Rajandream, MA,

Rogers, J, Rutter, S, Seeger, K, Skelton, J, Squares, R, Squares, S, Sulston, JE, Taylor, K, Whitehead, S, Barrell, BG 1998, 'Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence', *Nature*, vol. 393, no. 6685, pp. 537-544.

Coll, F, McNerney, R, Preston, MD, Guerra-Assunção, JA., Warry, A, Hill-Cawthorne, G, Mallard, K, Nair, M, Miranda, A, Alves, A, Perdigão, J, Viveiros, M, Portugal, I, Hasan, Z, Hasan, R, Glynn, JR, Martin, N, Pain, A 2015, 'Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences', *Genome Med*, vol. 7, no. 1, doi: 10.1186/s13073-015-0164-0.

Comas, I, Chakravartti, J, Small, PM, Galagan, J, Niemann, S, Kremer, K, Ernst, JD, Gagneux, S 2010, 'Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved', *Nat Genet*, vol. 42, no. 6, pp. 498-503.

Comas, I, Coscolla, M, Luo, T, Borrell, S, Holt, KE, Kato-Maeda, M, Parkhill, J, Malla, B, Berg, S, Thwaites, G, Yeboah-Manu, D, Bothamley, G, Mei, J, Wei, L, Bentley, S, Harris, SR, Niemann, S, Diel, R, Aseffa, A, Gao, Q, Young, D & Gagneux, S 2013, 'Out-of-Africa migration and Neolithic co-expansion of *Mycobacterium tuberculosis* with modern humans', *Nat Genet*, vol. 45, no. 10, doi: 10.1038/ng.2744.

Comas, I, Homolka, S. Niemann, S & Gagneux, S 2009, 'Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies', *PLoS One*, vol. 4, no. 11, doi: 10.1371/journal.pone.0007815.

Communicable Diseases Intelligence 2003, ‘Notifiable diseases surveillance, 1917 to 1991’, [Electronic edition] Retrieved from: http://www.health.gov.au/internet/main/publishing.nsf/Content/cda-pubs-annlrpt-oz_dis19_91.htm

Corbett, EL, Charalambous, S, Moloi, VM, Fielding, K, Grant, AD, Dye, C, De Cock, KM, Hayes, RJ, Williams, BG & Churchyard, GJ 2004, ‘Human immunodeficiency virus and the prevalence of undiagnosed tuberculosis in African gold miners’, *Am J Respir Crit Care Med*, vol. 170, no. 6, pp. 673-679.

Cousins, DV, Dawson, DJ 1999, ‘Tuberculosis due to Mycobacterium bovis in the Australian population: cases recorded during 1970-1994’, *Int J Tuberc Lung Dis*, vol. 3, no. 8, pp. 715-721.

Cousins, DV, Williams, SN, Dawson, DJ,1999, ‘Tuberculosis due to Mycobacterium bovis in the Australian population: DNA typing of isolates, 1970-1994’, *Int J Tuberc Lung Dis*, vol. 3, no.8, pp. 722-731.

Cuevas-Córdoba, B, Eusebio, DMJ, Almaraz-Velazco, R, Muñiz-Salazar, R, Laniado-Laborin, R & Zenteno-Cuevas, R 2015, ‘Mutation at embB 306, a potential marker for the identification of multidrug resistance associated with ethambutol in *Mycobacterium tuberculosis*’, *Antimicrob Agents and Chemother*, vol. 59, no. 9, pp. 5455-5462.

Dale, JW, Nor, RM, Ramayah, S, Tang, TH & Zainuddin, Z F 1999, ‘Molecular epidemiology of tuberculosis in Malaysia’, *J Clin Microbiol*, vol. 37, no. 5, pp. 1265-1268.

Davies, P, Gordon, S & Davies, G 2014, 'Clinical Tuberculosis', London: CRC Press.

Davis, RJ, Jensen, SO, Van Hal, S, Espedido, B, Gordon, A, Farhat, R & Chan, R 2015, 'Whole-genome Sequencing in Real-Time Investigation and Management of a Pseudomonas aeruginosa Outbreak on a Neonatal Intensive Care Unit', *Infect Control Hosp Epidemiol*, vol. 36, no. 9, pp.1058-1064.

de Beer, JL, van Ingen, J, de Vries, G, Erkens, C, Sebek, M, Mulder, A, Sloot, R, van den Brandt, AM, Enaimi, M, Kremer, K, Supply, P & van Soolingen, D 2013, 'Comparative Study of IS6110 Restriction Fragment Length Polymorphism and Variable-Number Tandem-Repeat Typing of *Mycobacterium tuberculosis* Isolates in the Netherlands, Based on a 5-Year Nationwide Survey', *J Clin Microbiol*, vol. 51, no. 4, pp. 1193-1198.

de Boer, AS, Borgdorff, MW, de Haas, PE, Nagelkerke, NJ, van Embden, JD & van Soolingen, D 1999, 'Analysis of rate of change of IS6110 RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates', *J Infect Dis*, vol. 180, no. 4, pp.1238-1244.

de Groote, MA, Gruppo, V, Woolhiser, LK, Orme, IM, Gilliland, JC & Lenaerts, AJ 2012, 'Importance of confirming data on the in vivo efficacy of novel antibacterial drug regimens against various strains of *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 56, no. 2, pp. 731-738.

de Zoysa, R, Shoemack, R, Vaughan, R & Vaughan, A 2001, 'A prolonged outbreak of tuberculosis in the North Island', *New Zealand Public Health Reports*, vol. 8, no. 1, pp. 1-3.

Demay, C, Liens, B, Burguière, T, Hill, V, Couvin, D, Millet, J, Mokrousov, I, Sola, C, Zozio, T & Rastogi, N, 2012, 'SITVITWEB-- A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology', *Infect Genet Evol*, vol. 12, no.4, pp. 755-766.

Denholm, J, Coulter, C, Bastian I & the National Tuberculosis Advisory Committee 2016, 'Defining a tuberculosis cluster or outbreak', *Commun Dis Intell Q Rep*, vol. 40, no. 3, pp. e356-359.

Desikan, P, 'Sputum smear microscopy in tuberculosis: Is it still relevant?', *Indian J Med Res*, vol. 137, no. 3, pp.442-444.

Desjardins, CA, Cohen, KA, Munsamy, V, Abeel, T, Maharaj, K, Walker, BJ, Shea, TP, Almeida, DV, Manson, AL, Salazar, A, Padayatchi, N, O'Donnell, MR, Mlisana, KP, Wortman, J, Birren, BW, Grosset, J, Earl, AM & Pym, A. S. 2016, 'Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate ald in D-cycloserine resistance', *Nat Genet*, vol. 48, no.5, pp. 544-551.

Dheda, K, Barry, CE & Maartens, G 2016, 'Tuberculosis', *Lancet*, vol. 387, no. 10024, pp.1211-1226.

Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999 2000, *Am J Respir Crit Care Med*, vol. 161, no. 4, pp. 1376-1395.

Didelot, X, Nell, S, Yang, I, Woltemate, S, van der Merwe, S & Suerbaum, S 2013, 'Genomic evolution and transmission of Helicobacter pyloriin two South African families', *Proc Natl Acad Sci USA*, vol. 110, no. 34, pp. 13880-13885.

Dobbins,C, Marishta, K, Kuehnert, P, Arbisi, M, Darnall, E, Conover, C, Howland, J, Powell, K, Althomsons, S, Bamrah, S, Garrett, D & Haddad 2012, '*M. Tuberculosis* outbreak associated with a homeless shelter-Kane County, Illinois, 2007-2011', [Electronic edition] Retrieved from: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6111a3.htm

Doughty, EL, Sergeant, MJ, Adetifa, I, Antonio, M & Pallen, MJ 2014, 'Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer', *PeerJ*, vol. 2, doi:10.7717/peerj.585.

Doyle K 2015, 'DNA Quantitation in Next-Generation Sequencing Library Workflows' [Electronic edition] Retrieved from: https://www.promega.com.au/resources/pubhub/dna-quantitation-in-next-generation-sequencing-library-workflows/

Drobniewski, FA & Pozniak, AL 1996, 'Molecular diagnosis, detection of drug resistance and epidemiology of tuberculosis', *Br J Hosp Med*, vol. 56, no. 5, pp. 204-208.

du Toit, LC, Pillay, V & Danckwerts, MP 2006, 'Tuberculosis chemotherapy: current drug delivery approaches'. *Respir Res*, vol. 7, no. 1, doi:10.1186/1465-9921-7-118.

Dunn, JJ, Starke, JR & Revell, PA 2016, 'Laboratory Diagnosis of *Mycobacterium tuberculosis* Infection and Disease in Children', *J Clin Microbiol*, vol.54, no. 6, pp.1434-1441.

Dutta, NK, Alsultan, A, Gniadek, TJ, Belchis, DA, Pinn, ML, Mdluli, KE, Nuermberger, EL, Peloquin, CA, & Karakousis, PC 2013, 'Potent Rifamycin-Sparing Regimen Cures Guinea Pig Tuberculosis as Rapidly as the Standard Regimen', *Antimicrob Agents Chemother*, vol. 57, no. 8, pp. 3910-3916.

Eamranond, P & Jaramillo, E 2001, 'Tuberculosis in children: reassessing the need for improved diagnosis in global control strategies', *Int J Tuberc Lung Dis*, vol. 5, no.7, pp. 594-603.

Eldholm, VMJ, Rieux, A, Lopez, B, Sobkowiak, B, Ritacco, V & Balloux F 2015, 'Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain', *Nat Commun*, vol. 6, no. 7119, doi: 10.1038/ncomms8119.

Ernst, JD, Trevejo-Nuñez, G & Banaiee, N 2007, 'Genomics and the evolution, pathogenesis, and diagnosis of tuberculosis', *J Clin Invest*, vol.117, no. 7, pp. 1738-1745.

Esmail, H, Barry, CE, Young, DB & Wilkinson, RJ 2014, 'The ongoing challenge of latent tuberculosis' *Philos Trans R Soc Lond B Biol Sci*, vol. 369, no.1645, doi:10.1098/rstb.2013.0437.

European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis 2006,

‘Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance’, *Emerg Infect Dis*, vol. 12, no. 5, pp. 736-743.

Feuerriegel, S, Schleusener, V, Beckert, P, Kohl, TA, Miotto, P, Cirillo, D M, Cabibbe, AM, Niemann, S & Fellenberg, K 2015, ‘PhyResSE: a Web Tool Delineating *Mycobacterium tuberculosis* Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data’, *J Clin Microbiol*, vol. 53, no. 6, pp.1908-1914.

Filliol, I, Motiwala, AS, Cavatore, M, Qi, W, Hazbón, MH, Bobadilla del Valle, M, Fyfe, J, García-García, L, Rastogi, N, Sola C, Zozio, T, Guerrero, MI, León, CI, Crabtree, J, Angiuoli, S, Eisenach, KD, Durmaz, R, Joloba, ML, Rendón, A, Sifuentes-Osornio, J, Ponce de León, A, Cave, MD, Fleischmann, R, Whittam, TS & Alland, D 2006, ‘Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set’, *J Bacteriol*, vol. 188, no. 2, pp. 759-772.

Firdessa, R, Berg, S, Hailu, E, Schelling, E, Gumi, B, Erenso, G, Gadisa, E, Kiros, T, Habtamu, M, Hussein, J, Zinsstag, J, Robertson, BD, Ameni, G, Lohan, AJ, Loftus, B, Comas, I, Gagneux, S, Tschopp, R, Yamuah, L, Hewinson, G, Gordon, SV, Young, DB & Aseffa, A 2013, ‘Mycobacterial Lineages Causing Pulmonary and Extrapulmonary Tuberculosis, Ethiopia’, *Emerg Infect Dis*, vol. 19, no. 3, pp. 460-463.

Fleischmann, RD, Alland, D, Eisen, JA, Carpenter, L, White, O, Peterson, J, DeBoy, R, Dodson, R, Gwinn, M, Haft, D, Hickey, E, Kolonay, JF, Nelson,

WC, Umayam, LA, Ermolaeva, M, Salzberg, SL, Delcher, A, Utterback, T, Weidman, J, Khouri, H, Gill, J, Mikula, A, Bishai, W, Jacobs, WR Jr, Venter, JC & Fraser, CM 2002, 'Whole-Genome Comparison of *Mycobacterium tuberculosis* Clinical and Laboratory Strains', *J Bacteriol*, vol.184, no. 19, pp. 5479-5490.

Forbes, BA 2017, 'Mycobacterial Taxonomy', *J Clin Microbiol*, vol. 55, no. 2, pp. 380-383.

Ford, N, Shubber, Z, Meintjes, G, Grinsztejn, B, Eholie, S, Mills, EJ, Davies, MA, Vitoria, M, Penazzato, M, Nsanzimana, S, Frigati, L, O'Brien, D, Ellman, T, Ajose, O, Calmy, A & Doherty, M 2015, 'Causes of hospital admission among people living with HIV worldwide: a systematic review and meta-analysis', *Lancet HIV*, vol. 2, no. 10, pp. e438-444.

Forrellad, MA, Klepp, LI, Gioffré, A, Sabio y García, J, Morbidoni, HR, Santangelo, MdlP, Cataldi AA & Bigi, F2013, 'Virulence factors of the *Mycobacterium tuberculosis* complex', *Virulence*, vol. 4, no. 1, pp. 3-66.

Foxman, B, Zhang, L, Koopman, JS, Manning, SD, & Marrs, CF 2005, 'Choosing an appropriate bacterial typing technique for epidemiologic studies', *Epidemiol Perspect Innov*, vol. 2, doi:10.1186/1742-5573-2-10.

Francis, JR, Manchikanti, P, Blyth, CC, Denholm, J, Lowbridge, C, Coulter, C, Donnan, E, Stapledon, R, Krause, VL & Waring, J 2018, 'Multidrug-resistant tuberculosis in Australia, 1998-2012', *Int J Tuberc Lung Dis*, vol. 22, no. 3, pp. 294-299.

Freimanis Hance, L, Steingart, KR, Hahn, CG, Pascopella, L, & Nolan, CM 2007, 'Field assessment of a model tuberculosis outbreak response plan for low-incidence areas', *BMC Public Health*, vol. 7, no.1, doi:10.1186/1471-2458-7-307.

Frieden, TR, Sterling, TR, Munsiff, SS, Watt, CJ, & Dye, C 2003, 'Tuberculosis', *Lancet*, vol. 362, no. 9387, pp. 887-899.

Friedland, G, Churchyard, GJ, & Nardell, E 2007, 'Tuberculosis and HIV Coinfection: Current State of Knowledge and Research Priorities', *J Infect Dis*, vol. 196 (supplement 1: s1-s3), doi:10.1086/518667.

Gagneux, S 2018, 'Ecology and evolution of *Mycobacterium tuberculosis*', *Nat Rev Microbiol*, vol. 16, no. 4, pp. 202-213.

Gagneux, S, DeRiemer, K, Van, T, Kato-Maeda, M, de Jong, BC, Narayanan, S, Nicol, M, Niemann, S, Kremer, K, Gutierrez, MC, Hilty, M, Hopewell, PC & Small, PM 2006, 'Variable host–pathogen compatibility in *Mycobacterium tuberculosis*', *Proc Natl Acad Sci USA*, vol. 103, no. 8, pp. 2869-2873.

Galperin, MY, Makarova, KS, Wolf, YI & Koonin, EV 2015, 'Expanded microbial genome coverage and improved protein family annotation in the COG database', *Nucleic Acids Res*, vol. 43 (Database issue: D261-D269), doi:10.1093/nar/gku1223.

Gan, M, Liu, Q, Yang, C, Gao, Q & Luo, T 2016, 'Deep Whole-Genome Sequencing to Detect Mixed Infection of *Mycobacterium tuberculosis*', *PLoS One*, vol. 11, no. 7, doi: 10.1371/journal.pone.0159029.

Ganguly, N, Siddiqui, I & Sharma, P 2008, 'Role of *M. tuberculosis* RD-1 region encoded secretory proteins in protective response and virulence', *Tuberculosis (Edinb)*, vol. 88, no. 6, pp. 510-517.

Gardy, JL, Johnston, JC, Ho Sui, SJ, Cook, VJ, Shah, L, Brodkin, E, Rempel, S, Moore, R, Zhao, Y, Holt, R, Varhol, R, Birol, I, Lem, M, Sharma, MK, Elwood, K, Jones, SJ, Brinkman, FS, Brunham, RC & Holt, R 2011, 'Whole-genome sequencing and social-network analysis of a tuberculosis outbreak', *N Engl J Med*, vol. 364, no. 8, pp. 730-739.

Gautam, SS, Haug, G, Cooley, L, Mac Aogáin, M & O'Toole, RF 2019, 'Intercontinental translocation of latent multi-drug resistant tuberculosis to Australia demonstrated by whole-genome sequencing', *Med J Aust*, vol. 210, no.5, doi: 10.5694/mja2.50044.

Gautam, SS, KC, R, Leong, KWC, Mac Aogáin, M & O'Toole, RF 2019a, 'A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens', *J Biol Methods*, vol. 6, no.1, doi: http://dx.doi.org/10.14440/jbm.2019.276.

Gautam, SS, Mac Aogáin, M & O'Toole, RF 2017, 'Draft Genome Sequence of the First Confirmed Isolate of Multidrug-Resistant *Mycobacterium tuberculosis* in Tasmania', *Genome Announc*, vol. 5, no. 44, doi:10.1128/genomeA.01230-17.

Gautam, SS, Mac Aogáin, M, Bower, JE, Basu, I & O'Toole, RF 2017a, 'Differential carriage of virulence-associated loci in the New Zealand Rangipo outbreak strain of *Mycobacterium tuberculosis*', *Infect Dis*, vol. 49, no. 9, pp. 680-688.

Gautam, SS, Mac Aogáin, M, Cooley, LA, Haug, G, Fyfe, JA, Globan, M & O'Toole, RF 2018, 'Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case', *PLoS One*, vol. 13, no. 2, doi: 10.1371/journal.pone.0192351

Geng, E, Kreiswirth, B, Driver, C, Li, J, Burzynski, J, DellaLatta, P, LaPaz, A & Schluger, NW 2002, 'Changes in the Transmission of Tuberculosis in New York City from 1990 to 1999', *N Engl J Med*, vol. 346, no. 19, pp. 1453-1458.

Gierke R, McGee L, Bealla B and Pilishivili T 2014, 'Manual for the Surveillance of Vaccine-Preventable Diseases', [Electronic edition] Retrieved from: http://www.cdc.gov/vaccines/pubs/surv-manual/chpt11-pneumo.html

Global Laboratory Initiative 2014, 'Mycobacteriology Laboratory Manual', [Electronic edition] Retrieved from: http://www.who.int/tb/laboratory/mycobacteriology-laboratory-manual.pdf

Gormus, BJ, Blanchard, JL, Alvarez, XH & Didier, PJ 2004, 'Evidence for a rhesus monkey model of asymptomatic tuberculosis', *J Med Primatol*, vol. 33, no. 3, pp.134-145.

Grad, YH, Lipsitch, M, Feldgarden, M, Arachchi, HM, Cerqueira, GC, FitzGerald, M, Godfrey, P, Haas, BJ, Murphy, CI, Russ, C, Sykes, S, Walker, BJ, Wortman, JR, Young, S, Zeng, Q, Abouelleil, A, Bochicchio, J, Chauvin, S, Desmet, T, Gujja, S, McCowan C, Montmayeur, A, Steelman, S, Frimodt-Møller, J, Petersen, AM, Struve, C, Krogfelt, KA, Bingen, E, Weill, FX, Lander, ES, Nusbaum, C, Birren, BW, Hung, DT & Hanage, WP

2012, ‘Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011’, *Proc Natl Acad Sci USA*, vol. 109, no. 8, pp. 3065-3070.

Guindon, S & Gascuel, O 2003, ‘A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood’, *Syst Biol*, vol. 52, no. 5, pp. 696-704.

Guindon, S, Delsuc, F, Dufayard, JF & Gascuel O, 2009, ‘Estimating maximum likelihood phylogenies with PhyML’, *Methods Mol Biol,* vol. 537, pp.113-37.

Gurjav, U, Outhred, AC, Jelfs, P, McCallum, N, Wang, Q, Hill-Cawthorne, GA, Marais BJ & Sintchenko, V 2016, ‘Whole-genome Sequencing Demonstrates Limited Transmission within Identified *Mycobacterium tuberculosis* Clusters in New South Wales, Australia’, *PLoS One*, vol. 11, no. 10, doi: 10.1371/journal.pone.0163612.

Gutacker, MM, Smoot, JC, Migliaccio, CA, Ricklefs, SM, Hua, S, Cousins, DV, Graviss, EA, Shashkina, E, Kreiswirth, BN & Musser, JM 2002, ‘Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains’, *Genetics*, vol. 162, no. 4, pp. 1533-1543.

Guthrie, JL, Delli Pizzi, A, Roth, D, Kong, C, Jorgensen, D, Rodrigues, M, Tang, P, Cook, VJ, Johnston J & Gardy, JL 2018, ‘Genotyping and Whole-Genome Sequencing to Identify Tuberculosis Transmission to Pediatric Patients in British Columbia, Canada, 2005–2014’, *J Infect Dis*, vol. 218, no. 7, pp. 1155-1163.

Gygli, SM, Borrell, S, Trauner, A & Gagneux, S 2017, 'Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives', *FEMS Microbiol Rev*, vol. 41, no. 3, pp. 354-373.

Gyles, C 2008, 'The DNA revolution', *Cand Vet J*, vol. 49, no. 8, pp. 745-746.

Halachev, MR, Chan, JZ, Constantinidou, CI, Cumley, N, Bradley, C, Smith-Banks, M, Oppenheim MJ & Pallen, MJ 2014, 'Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England', *Genome Med*, vol. 6, no. 11, doi:10.1186/s13073-014-0070-x.

Han, SJ, Song, T, Cho, YJ, Kim, JS, Choi, SY, Bang, HE, Chun, J, Bai, GH, Cho, SN & Shin, SJ 2015, 'Complete genome sequence of *Mycobacterium tuberculosis* K from a Korean high school outbreak, belonging to the Beijing family', *Stand Genomic Sci*, vol. 10, doi:10.1186/s40793-015-0071-4

Hanway A, Comiskey, C, Tobin, K, O'Toole, RF 2016, 'Relating annual migration from high tuberculosis burden country of origin to changes in foreign-born tuberculosis notification rates in low-medium incidence European countries', *Tuberculosis (Edinb)*, vol.101, pp. 67-74.

Harris, SR, Cartwright, EJP, Török, ME, Holden, MTG, Brown, NM, Ogilvy-Stuart, AL, Ellington, MJ, Quail, MA, Bentley, SD, Parkhill, J & Peacock, SJ 2013, 'Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study', *Lancet Infect Dis*, vol. 13, no. 2, pp. 130-136.

Hartkoorn, RC, Uplekar, S & Cole, ST 2014, 'Cross-Resistance between Clofazimine and Bedaquiline through Upregulation of MmpL5 in *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 58, no. 5, pp. 2979-2981.

Hatherell H-A, Colijn, C, Stagg, HR, Jackson, C, Winter, JR & Abubakar 2016, 'Interpreting whole-genome sequencing for investigating tuberculosis transmission: a systematic review', *BMC Med*, vol. 14, no. 21, doi: 10.1186/s12916-016-0566-x.

Haver, HL, Chua, A, Ghode, P, Lakshminarayana, SB, Singhal, A, Mathema, B, Wintjens R & Bifani, P 2015, 'Mutations in Genes for the F420 Biosynthetic Pathway and a Nitroreductase Enzyme Are the Primary Resistance Determinants in Spontaneous In Vitro Selected PA-824-Resistant Mutants of *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 59, no. 9, pp. 5316-5323.

Healey, A, Furtado, A, Cooper, T & Henry, RJ 2014, 'Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species', *Plant Methods*, vol. 10, no. 1, pp. 1-8.

Hett, EC & Rubin, EJ 2008, 'Bacterial growth and cell division: a mycobacterial perspective', *Microbiol Mol Biol Rev*, vol. 72, no. 1, pp. 126-156.

Heym, B, Alzari, PM, Honore, N & Cole, ST 1995, 'Missense mutations in the catalase-peroxidase gene, katG, are associated with isoniazid resistance in *Mycobacterium tuberculosis*', *Mol Microbiol*, vol. 15, no. 2, pp. 235-245.

Hillemann, D, Rüsch-Gerdes, S & Richter, E 2008, 'In vitro-selected linezolid-resistant *Mycobacterium tuberculosis* mutants', *Antimicrob Agents Chemother*, vol. 52, no. 2, pp. 800-801.

Houben, RMGJ & Dodd, PJ 2016, 'The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling', PLoS Med, vol. 13, no. 10, doi: 10.1371/journal.pmed.1002152.

Huitric, E, Verhasselt, P, Koul, A, Andries, K, Hoffner, S & Andersson, DI 2010, 'Rates and mechanisms of resistance development in *Mycobacterium tuberculosis* to a novel diarylquinoline ATP synthase inhibitor', *Antimicrob Agents Chemother*, vol. 54, no. 3, pp. 1022-1028

Ilina, EN, Shitikov, EA, Ikryannikova, LN, Alekseev, DG, Kamashev, DE, Malakhova, MV, Parfenova, TV, Afanas'ev, MV, Ischenko, DS, Bazaleev, NA, Smirnova, TG, Larionova, EE, Chernousova, LN, Beletsky, AV, Mardanov, AV, Ravin, NV, Skryabin, KG & Govorun, VM 2013, 'Comparative genomic analysis of Mycobacterium tuberculosis drug resistant strains from Russia', *PLoS One, vol.* 8, no. 2, doi: 10.1371/journal.pone.0056577.

Illumina Inc. 2014, Nextera® Library validation and cluster density optimization', [Electronic edition] Retrieved from: https://www.illumina.com/documents/products/technotes/technote_nextera_library_validation.pdf

Illumina, Inc. 2012, 'Nextera® XT DNA Sample Preparation Guide', [Electronic edition] Retrieved from:

http://www.liai.org/files/nextera_xt_sample_preparation_guide_15031942_c.pdf

Institute of Environmental Science and Research Ltd 2015, 'Tuberculosis in New Zealand: Annual Report 2014', [Electronic edition] Retrieved from: https://surv.esr.cri.nz/surveillance/AnnualTBReports.php?we_objectID=4251

International Human Genome Sequencing Consortium 2004, 'Finishing the euchromatic sequence of the human genome', *Nature*, vol. 431, no. 7011, pp. 931-945.

Ioerger, TR, Feng, Y, Ganesula, K, Chen, X, Dobos, KM, Fortune, S, Jacobs, WR Jr, Mizrahi, V, Parish, T, Rubin, E, Sassetti, C & Sacchettini, JC 2010, 'Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories', *J Bacteriol*, vol. 192, no. 14, pp. 3645-3653.

Iwamoto, T, Fujiyama, R, Yoshida, S, Wada, T, Shirai, C & Kawakami, Y 2009, 'Population structure dynamics of *Mycobacterium tuberculosis* Beijing strains during past decades in Japan', *J Clin Microbiol*, vol. 47, no. 10, pp. 3340-3343.

Jajou, R, Neeling, Ad, Hunen, Rv, Vries, Gd, Schimmel, H, Mulder, A, Anthony, R, Hoek, WvdH & Soolingen, DV 2018, 'Epidemiological links between tuberculosis cases identified twice as efficiently by whole-genome sequencing than conventional molecular typing: A population-based study', *PLoS One*, vol. 13, no. 4, doi: 10.1371/journal.pone.0195413

Jones, BE, Young, SM, Antoniskis, D, Davidson, PT, Kramer, F & Barnes, PF 1993, 'Relationship of the manifestations of tuberculosis to CD4 cell counts in patients with human immunodeficiency virus infection', *Am Rev Respir Dis*, vol. 148, no. 5, pp. 1292-1297.

Joshi, JM 2011, 'Tuberculosis chemotherapy in the 21(st) century: Back to the basics', *Lung India*, vol. 28, no. 3, pp.193-200.

Kambli, P, Ajbani, K, Nikam, C, Sadani, M, Shetty, A, Udwadia, Z, Georghiou, SB, Rodwell, TC, Catanzaro, A & Rodrigues, C 2016, 'Correlating rrs and eis promoter mutations in clinical isolates of *Mycobacterium tuberculosis* with phenotypic susceptibility levels to the second-line injectables', *Int J Mycobacteriol*, vol 5, no. 1, pp. 1-6.

Kamerbeek, J, Schouls, L, Kolk, A, van Agterveld, M, van Soolingen, D, Kuijper, S, Bunschoten, A, Molhuizen, H, Shaw, R, Goyal, M & van Embden, J 1997, 'Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology', *J Clin Microbiol*, vol. 35, no. 5, pp. 907-914.

Kamper-Jorgensen, Z, Andersen, AB, Kok-Jensen, A, Kamper-Jorgensen, M, Bygbjerg, IC, Andersen, PH, Thomsen VO & Lillebaek T 2012, 'Migrant tuberculosis: the extent of transmission in a low burden country', *BMC Infect Dis*, vol. 12, no. 60, doi: 10.1186/1471-2334-12-60.

Kanehisa, M & Goto, S 2000, 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Res*, vol. 28, no. 1, pp. 27-30.

Kanehisa, M, Furumichi, M, Tanabe, M, Sato, Y & Morishima, K 2016, 'KEGG: new perspectives on genomes, pathways, diseases and drugs', *Nucleic Acids Res*, vol. 45, doi:10.1093/nar/gkw1092

Kaplan, G 2005, 'Rational Vaccine Development-A New Trend in Tuberculosis Control', *N Engl J Med*, vol. 353, no. 15, pp.1624-1625.

Kato-Maeda, M & Nahid, P 2012, '*Mycobacterium tuberculosis* Lineage-What's in Your Lungs?', *Clin Infect Dis*, vol. 54, no. 2, pp. 220-224.

Kato-Maeda, M, Ho, C, Passarelli, B, Banaei, N, Grinsdale, J, Flores, L, Anderson, J, Murray, M, Rose, G, Kawamura, LM, Pourmand, N, Tariq, MA, Gagneux S & Hopewell, PC 2013, 'Use of Whole-genome Sequencing to Determine the Microevolution of *Mycobacterium tuberculosis* during an Outbreak', *PLoS One*, vol. 8, no. 3, doi: 10.1371/journal.pone.0058235

Kato-Maeda, M, Metcalfe, JZ & Flores, L 2011, 'Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies', *Future Microbiol*, vol. 6, no. 2, pp. 203-216.

Kato-Maeda, M, Rhee, JT, Gingeras, TR, Salamon, H, Drenkow, J, Smittipat, N & Small, PM 2001, 'Comparing genomes within the species *Mycobacterium tuberculosis*', *Genome Res*, vol. 11, no. 4, pp. 547-554.

Kearse, M, Moir, R, Wilson, A, Stones-Havas, S, Cheung, M, Sturrock, S, Buxton, S, Cooper, A, Markowitz, S, Duran, C, Thierer, T, Ashton, B, Meintjes, P & Drummond, A 2012, 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, vol. 28, no. 12, pp. 1647-1649.

Kirigia, JM & Muthuri, RDK 2016, 'Productivity losses associated with tuberculosis deaths in the World Health Organization African region', *Infect Dis Poverty*, vol. 5, doi:10.1186/s40249-016-0138-5

Knechel, NA 2009, 'Tuberculosis: Pathophysiology, Clinical Features, and Diagnosis', *Crit Care Nurse*, vol. 29, no. 2, pp. 34-43.

Koch, A & Mizrahi, V 2018, '*Mycobacterium tuberculosis*', *Trends in Microbiol*, vol. 26, no. 6, pp. 555-556.

Koch, R 1882, 'The etiology of tuberculosis', American Society for Microbiology, [Electronic edition] Retrieved from: https://www.asm.org/ccLibraryFiles/FILENAME/0000000228/1882p109.pdf

Kochi, A, Vareldzis, B & Styblo, K 1993, 'Multidrug-resistant tuberculosis and its control', *Res Microbiol*, vol. 144, no. 2, pp. 104-110.

Konno, K, Feldmann, FM & McDermott, W 1967, 'Pyrazinamide susceptibility and amidase activity of tubercle bacilli', *Am Rev Respir Dis*, vol. 95, no. 3, pp. 461-469.

Köser, CU, Ellington, MJ, Cartwright, EJ, Gillespie, SH, Brown, NM, Farrington, M, Holden, MT, Dougan, G, Bentley, SD, Parkhill, J & Peacock, SJ 2012, 'Routine use of microbial whole-genome sequencing in diagnostic and public health microbiology', *PLoS Pathog*, vol. 8, no. 8, doi: 10.1371/journal.ppat.1002824

Köser, CU, Holden, MTG, Ellington, MJ, Cartwright, EJP, Brown, NM., Ogilvy-Stuart, AL, Hsu, LY, Chewapreecha, C, Croucher, NJ, Harris, SR, Sanders, M, Enright, MC, Dougan, G, Bentley, SD, Parkhill, J, Fraser, LJ,

Betley, JR, Schulz-Trieglaff, OB, Smith, GP & Peacock, SJ 2012, 'Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak', *N Engl J Med*, vol. 366, no. 24, pp. 2267-2275.

Kremer, K, Glynn, JR, Lillebaek, T, Niemann, S, Kurepina, NE, Kreiswirth, BN, Bifani PJ & van Soolingen, D 2004, 'Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers', *J Clin Microbiol*, vol. 42, no. 9, pp. 4040-4049.

Kubica, GP, Kim, TH & Dunbar FP 1972, 'Designation of Strain H37Rv as the Neotype of *Mycobacterium tuberculosis*', *Int J Syst Evol Microbiol*, vol. 22, no. 2, pp. 99-106.

Kulchavenya, E 2014, 'Extrapulmonary tuberculosis: are statistical reports accurate?', *Ther Adv Infect Dis*, vol. 2, no. 2, pp. 61-70.

Kumar, N, Bhargava, SK, Agrawal, CS, George, K, Karki, P, & Baral, D 2005, 'Chest radiographs and their reliability in the diagnosis of tuberculosis', *J Nepal Med Assoc*, vol. 44, no. 160, pp. 138-142.

Kwan, CK & Ernst, JD 2011, 'HIV and Tuberculosis: A Deadly Human Syndemic', *Clin Microbiol Rev*, vol. 24, no. 2, pp. 351-376.

Lalor, MK., Casali, N, Walker, TM, Anderson, LF, Davidson, JA, Ratna, N, Mullarkey, C, Gent, M, Foster, K, Brown, T, Magee, J, Barrett, A, Crook, DW, Drobniewski, F, Thomas, HL & Abubakar, I 2018, 'The use of whole-genome sequencing in cluster investigation of a multidrug-resistant tuberculosis outbreak', *Eur Respir J*, vol. 51, no. 6, doi:10.1183/13993003.02313-2017.

Lecuit, M & Eloit, M 2014, 'The diagnosis of infectious diseases by whole-genome next generation sequencing: a new era is opening', *Front Cell Infect Microbiol*, vol. 4, doi:10.3389/fcimb.2014.00025.

Lee, MP, Chan, JWM, Ng, KK, & Li, PC 2000, 'Clinical manifestations of tuberculosis in HIV-infected patients', *Respirology*, vol. 5, no. 4, pp. 423-426.

Lee, RS & Pai, M 2017, 'Real-Time Sequencing of *Mycobacterium tuberculosis*: Are We There Yet?', *J Clin Microbiol*, vol. 55, no. 5, pp. 1249-1254.

Lee, SH 2016, 'Tuberculosis Infection and Latent Tuberculosis', *Tuber Respir Dis*, vol. 79, no. 4, pp. 201-206.

Leong, KWC, Cooley, LA, Anderson, TL, Gautam, SS, McEwan, B, Wells, A, Wilson, F, Hughson, L & O'Toole, RF 2018, 'Emergence of Vancomycin-Resistant Enterococcus faecium at an Australian Hospital: A Whole-genome Sequencing Analysis', *Sci Rep*, vol. 8, no. 1, doi: 10.1038/s41598-018-24614-6

Li, H & Durbin, R 2010, 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*, vol. 26, no. 5, pp. 589-595.

Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, Marth, G, Abecasis, G, Durbin, R & 1000 Genome Project Data Processing Subgroup 2009, 'The sequence alignment/Map format and SAMtools', *Bioinformatics*, vol. 25, no. 16, pp. 2078-2079.

Liang, L, Wu, Q, Gao, L, Hao, Y, Liu, C, Xie, Y, Sun, H, Yan, X, Li, F, Li, H, Fang, H, Ning, N, Cui, Y & Han, L 2012, 'Factors contributing to the high prevalence of multidrug-resistant tuberculosis: a study from China', *Thorax*, vol. 67, no. 7, pp. 632-638.

Liu, PI, McGregor, DH, Faucher, I, Jinks, WL, Miller, LA, Green, L & Liu, JG 1973, 'Comparison of Three Culture Media for Isolation of *Mycobacterium tuberculosis*: a 6-Year Study', *Appl Microbiol*, vol. 26, no. 6, pp. 880-883.

Liu, X, Gutacker, MM, Musser, JM & Fu, YX 2006, 'Evidence for Recombination in *Mycobacterium tuberculosis*', *J Bacteriol*, vol. 188, no. 23, pp. 8169-8177.

Llarena, AK, Taboada, E & Rossi, M 2017, 'Whole-Genome Sequencing in Epidemiology of Campylobacter jejuni Infections', *J Clin Microbiol*, vol. 55, no. 5, pp. 1269-1275.

Lyu, J, Lee, SG, Hwang, S, Lee, SO, Cho, OH, Chae, EJ, Lee, SD, Kim, WS, Kim, DS & Shim, TS 2011, 'Chest computed tomography is more likely to show latent tuberculosis foci than simple chest radiography in liver transplant candidates', *Liver Transpl*, vol. 17, no. 8, pp. 963-968.

Mac Aogain, M, Gautam, SS, Bower, JE, Basu, I & O'Toole, RF2016, 'Draft Genome Sequence of a New Zealand Rangipo Strain of *Mycobacterium tuberculosis*', *Genome Announc*, vol. 4, no. 4, doi:10.1128/genomeA.00657-16

MacCannell, D 2013, 'Bacterial strain typing', *Clin Lab Med*, vol. 33, no. 3, pp. 629-650.

Malla, B, Stucki, D, Borrell, S, Feldmann, J, Maharjan, B, Shrestha, B, Fenner, L & Gagneux S 2012, 'First insights into the phylogenetic diversity of *Mycobacterium tuberculosis* in Nepal', *PLoS One*, vol. 7, no. 12, doi: 10.1371/journal.pone.0052297.

Mangtani, P, Abubakar, I, Ariti, C, Beynon, R, Pimpin, L, Fine, PE, Rodrigues, LC, Smith, PG, Lipman, M, Whiting, PF & Sterne, JA 2014, 'Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials', *Clin Infect Dis*, vol. 58, no. 4, pp. 470-480.

Manson, AL, Abeel, T, Galagan, JE, Sundaramurthi, JC, Salazar, A, Gehrmann, T, Shanmugam, SK, Palaniyandi, K, Narayanan, S, Swaminathan, S & Earl, AM 2017, '*Mycobacterium tuberculosis* Whole-genome Sequences from Southern India Suggest Novel Resistance Mechanisms and the Need for Region-Specific Diagnostics', *Clin Infect Dis*, vol. 64, no. 11, pp.1494-1501.

Marais, BJ, Gie, RP, Schaaf, HS, Beyers, N, Donald, PR & Starke, JR 2006, 'Childhood pulmonary tuberculosis: old wisdom and new challenges', *Am J Respir Crit Care Med*, vol. 173, no. 10, pp. 1078-1090.

Marquina-Castillo, B, García-García, L, Ponce-de-León, A, Jimenez-Corona, ME, Bobadilla-del Valle, M, Cano-Arellano, B, Canizales-Quintero, S, Martinez-Gamboa, A, Kato-Maeda, M, Robertson, B, Young, D, Small P, Schoolnik, G, Sifuentes-Osornio, J & Hernandez-Pando, R 2009, 'Virulence, immunopathology and transmissibility of selected strains of *Mycobacterium tuberculosis* in a murine model', *Immunology*, vol. 128, no. 1, pp. 123-133, doi:10.1111/j.1365-2567.2008. 03004.x.

Maryandyshev, A, Pontali, E, Tiberi, S, Akkerman, O, Ganatra, S, Sadutshang, TD, Alffenaar, JW, Amale, R, Mullerpattan, J, Topgyal, S, Udwadia, ZF, Centis, R, D'Ambrosio, L, Sotgiu, G & Migliori, GB 2017, 'Bedaquiline and Delamanid Combination Treatment of 5 Patients with Pulmonary Extensively Drug-Resistant Tuberculosis', *Emerg Infect Dis*, vol. 23, no.10, doi: 10.3201/eid2310.170834.

Maus, CE, Plikaytis, BB & Shinnick, TM 2005, 'Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 49, no. 8, pp. 3192-3197.

Maus, CE, Plikaytis, BB & Shinnick, TM 2005, 'Mutation of tlyA confers capreomycin resistance in *Mycobacterium tuberculosis*' *Antimicrob Agents Chemother*, vol. 49, no. 2, pp. 571-577.

Mazars, E, Lesjean, S, Banuls, AL, Gilbert, M, Vincent, V, Gicquel, B, Tibayrenc, M, Locht, C & Supply, P 2001, 'High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology', *Proc Natl Acad Sci USA*, vol. 98, no. 4, pp. 1901-1906.

McElnay, C, Thornley, C & Armstrong, R 2004, 'A community and workplace outbreak of tuberculosis in Hawke's Bay in 2002', *N Z Med J*, vol. 117, no. 1200, U1019.

McGinnis, J, Laplante, J, Shudt, M & George, KS 2016, 'Next generation sequencing for whole-genome analysis and surveillance of influenza A viruses', *J Clin Virol*, 79, pp. 44-50.

McLernon, J, Costello, E, Flynn, O, Madigan, G & Ryan, F 2010, 'Evaluation of Mycobacterial Interspersed Repetitive-Unit-Variable-Number Tandem-Repeat Analysis and Spoligotyping for Genotyping of Mycobacterium bovis Isolates and a Comparison with Restriction Fragment Length Polymorphism Typing', *J Clin Microbiol*, vol. 48, no. 12, pp. 4541-4545.

Meier, A, Kirschner, P, Bange, FC, Vogel, U & Böttger, E 1994, 'Genetic alterations in streptomycin-resistant *Mycobacterium tuberculosis*: mapping of mutations conferring resistance', *Antimicrob Agents Chemother*, vol. 38, no. 2, pp. 228-233.

Mestre, O, Luo, T, Dos Vultos, T, Kremer, K, Murray, A, Namouchi, A, Jackson, C, Rauzier, J, Bifani, P, Warren, R, Rasolofo, V, Mei, J, Gao, Q & Gicquel, B 2011, 'Phylogeny of *Mycobacterium tuberculosis* Beijing Strains Constructed from Polymorphisms in Genes Involved in DNA Replication, Recombination and Repair', *PLoS One*, vol. 6, no.1, doi: 10.1371/journal.pone.0016020.

Metsky, HC, Matranga, CB, Wohl, S, Schaffner, SF, Freije, CA, Winnicki, SM, West, K, Qu J, Baniecki, ML, Gladden-Young, A, Lin, AE, Tomkins-Tinch, CH, Ye, SH, Park, DJ, Luo, CY, Barnes, KG, Shah, RR, Chak, B, Barbosa-Lima, G, Delatorre, E, Vieira, YR, Paul, LM, Tan, AL, Barcellona, CM, Porcelli, MC, Vasquez, C, Cannons, AC, Cone, MR, Hogan, KN, Kopp, EW, Anzinger, JJ, Garcia, KF, Parham, LA, Ramírez, RMG, Montoya, MCM, Rojas, DP, Brown, CM, Hennigan, S, Sabina, B, Scotland, S, Gangavarapu, K, Grubaugh, ND, Oliveira, G, Robles-Sikisaka, R, Rambaut, A, Gehrke, L, Smole, S, Halloran, ME, Villar, L, Mattar, S, Lorenzana, I, Cerbino-Neto, J,

Valim, C, Degrave, W, Bozza, PT, Gnirke, A, Andersen, KG, Isern, S, Michael, SF, Bozza, FA, Souza, TML, Bosch, I, Yozwiak, NL, MacInnis, BL & Sabeti, PC 2017, 'Zika virus evolution and spread in the Americas', *Nature*, vol. 546, no. 7658, pp. 411-415.

Migliori, GB, Lange, C, Girardi, E, Centis, R, Besozzi, G, Kliiman, K, Ortmann, J, Matteelli, A, Spanevello, A, Cirillo DM & SMIRA/TBNET Study Group 2008, 'Extensively drug-resistant tuberculosis is worse than multidrug-resistant tuberculosis: different methodology and settings, same results', *Clin Infect Dis*, vol. 46, no. 6, pp. 958-959.

Miotto, P, Cabibbe, AM, Feuerriegel, S, Casali, N, Drobniewski, F, Rodionova, Y, Bakonyte, D, Stakenas, P, Pimkina, E, Augustynowicz-Kopeć, E, Degano, M, Ambrosi, A, Hoffner, S, Mansjö, M, Werngren, J, Rüsch-Gerdes, S, Niemann S & Cirillo DM 2014, '*Mycobacterium tuberculosis* pyrazinamide resistance determinants: a multicenter study', *MBio*, vol. 5, no. 5, doi: 10.1128/mBio.01819-14.

Molle, V, Reynolds, Robert C, Alderwick, Luke J, Besra, Gurdyal S, Cozzone, Alain J, Fütterer, K, & Kremer, L 2008, 'EmbR2, a structural homologue of EmbR, inhibits the *Mycobacterium tuberculosis* kinase/substrate pair PknH/EmbR', *Biochem J*, vol. 410, no. 2, pp. 309-317.

More, SJ, Radunz B & Glanville, RJ 2015, 'Lessons learned during the successful eradication of bovine tuberculosis from Australia', *Vet Rec*, vol. 117, no. 9, pp. 224-232.

Morlock, GP, Metchock, B, Sikes, D, Crawford, JT & Cooksey, RC 2003, 'ethA, inhA, and katG loci of ethionamide-resistant clinical *Mycobacterium*

tuberculosis isolates', *Antimicrob Agents Chemother*, vol. 47, no.12, pp. 3799-3805.

Mostowy, S, Cousins, D, Brinkman, J, Aranaz, A & Behr, MA 2002, 'Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex', *J Infect Dis*, vol. 186, no. 1, pp. 74-80.

Munoz-Elias, EJ & McKinney, JD 2005, '*Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence', *Nat Med*, vol. 11, no. 6, pp. 638-644.

Namasivayam, S, Maiga, M, Yuan, W, Thovarai, V, Costa, DL, Mittereder, LR, Wipperman, MF, Glickman, MS, Dzutsev, A, Trinchieri, G & Sher, A 2017, 'Longitudinal profiling reveals a persistent intestinal dysbiosis triggered by conventional anti-tuberculosis therapy', *Microbiome*, vol. 5, no. 1, doi:10.1186/s40168-017-0286-2.

Nebenzahl-Guimaraes, H, Yimer, SA, Holm-Hansen, C, de Beer, J, Brosch, R & van Soolingen, D 2016, 'Genomic characterization of *Mycobacterium tuberculosis* lineage 7 and a proposed name: 'Aethiops vetus', *Microb Genom*, vol. 2, no. 6, doi: 10.1099/mgen.0.000063.

Nikolayevskyy VKK, Niemann S & Drobniewski F 2016, 'Whole-genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: A systematic review', *Tuberculosis (Edinb)*, vol. 98, pp. 77-85.

Norbis, L, Alagna, R, Tortoli, E, Codecasa, LR, Migliori, GB & Cirillo, DM 2014, 'Challenges and perspectives in the diagnosis of extrapulmonary tuberculosis', *Expert Rev Anti Infect Ther*, vol. 12, no. 5, pp. 633-647.

Nübel, U, Dordel, J, Kurt, K, Strommenger, B, Westh, H, Shukla, SK, Zemlicková, H, Leblois, R, Wirth, T, Jombart, T, Balloux, F, Witte, W 2010, 'A Timescale for Evolution, Population Expansion, and Spatial Spread of an Emerging Clone of Methicillin-Resistant Staphylococcus aureus', *PLoS Pathog*, vol. 6, no. 4, doi: 10.1371/journal.ppat.1000855.

Okumura, K, Kato, M, Kirikae, T, Kayano, M & Miyoshi-Akiyama, T 2015, 'Construction of a virtual *Mycobacterium tuberculosis* consensus genome and its application to data from a next generation sequencer', *BMC Genomics*, vol. 16, no. 1, doi: 10.1186/s12864-015-1368-9.

O'Toole, RF & Gautam, SS 2017, 'Limitations of the *Mycobacterium tuberculosis* reference genome H37Rv in the detection of virulence-related loci', *Genomics*, vol. 109, no. 5, pp. 471-474.

Pai, M, Behr, MA, Dowdy, D, Dheda, K, Divangahi, M, Boehme, CC, Ginsberg, A, Swaminathan, S, Spigelman, M, Getahun, H, Menzies, D & Raviglione, M 2016, 'Tuberculosis', *Nat Rev Dis Primers*, vol. 27, no. 2, doi: 10.1038/nrdp.2016.76.

Pankhurst, LJ, Del Ojo Elias, C, Votintseva, AA, Walker, TM, Cole, K, Davies, J, Fermont, JM, Gascoyne-Binzi, DM, Kohl, TA, Kong, C, Lemaitre, N, Niemann, S, Paul, J, Rogers, TR, Roycroft, E, Smith, EG, Supply, P, Tang, P, Wilcox, MH, Wordsworth, S, Wyllie, D, Xu, L, Crook, DW & COMPASS-TB Study Group. 2016, 'Rapid, comprehensive, and affordable

mycobacterial diagnosis with whole-genome sequencing: a prospective study', *Lancet Respir Med*, vol. 4, no. 1, pp. 49-58.

Papavinasasundaram, KG, Chan, B, Chung, JH, Colston, MJ, Davis, EO & Av-Gay, Y 2005 'Deletion of the *Mycobacterium tuberculosis* pknH Gene Confers a Higher Bacillary Load during the Chronic Phase of Infection in BALB/c Mice', *J Bacteriol*, vol. 187, no. 16, pp. 5751-5760.

Park, DJ, Dudas, G, Wohl, S., Goba, A., Whitmer, S. L., Andersen, KG, Sealfon, RS, Ladner, JT, Kugelman, JR, Matranga, CB, Winnicki, SM, Qu, J, Gire, SK, Gladden-Young, A, Jalloh, S, Nosamiefan, D, Yozwiak, NL, Moses, LM, Jiang, PP, Linm AE, Schaffner, SF, Bird, B, Towner, J, Mamoh, M, Gbakie, M, Kanneh, L, Kargbo, D, Massally, JL, Kamara, FK, Konuwa, E, Sellu, J, Jalloh, AA, Mustapha, I, Foday, M, Yillah, M, Erickson, BR, Sealy, T, Blau, D, Paddock, C, Brault, A, Amman, B, Basile, J, Bearden, S, Belser, J, Bergeron, E, Campbell, S, Chakrabarti, A, Dodd, K, Flint, M, Gibbons, A, Goodman, C, Klena, J, McMullan, L, Morgan, L, Russell, B, Salzer, J, Sanchez, A, Wang, D, Jungreis, I, Tomkins-Tinch, C, Kislyuk, A, Lin, MF, Chapman, S, MacInnis, B, Matthews, A, Bochicchio, J, Hensley, LE, Kuhn, JH, Nusbaum, C, Schieffelin, JS, Birren, BW, Forget, M, Nichol, ST, Palacios, GF, Ndiaye, D, Happi, C, Gevao, SM, Vandi, MA,, Kargbo, B, Holmes, EC, Bedford, T, Gnirke, A, Ströher, U, Rambaut, A, Garry, RF & Sabeti, PC 2015, 'Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone', *Cell*, vol. 161, no. 7, pp. 1516-1526.

Parkinson, NJ, Maslau, S, Ferneyhough, B, Zhang, G, Gregory, L, Buck, D, Ragoussis, J, Ponting, CP & Fischer, MD 2012, 'Preparation of high-quality

next-generation sequencing libraries from picogram quantities of target DNA', *Genome Res*, vol. 22, no. 1, pp. 125-133.

Parwati, I, Alisjahbana, B, Apriani, L, Soetikno, RD, Ottenhoff, TH, van der Zanden, AG, van der Meer, J, van Soolingen, D & van Crevel, R 2010, '*Mycobacterium tuberculosis* Beijing genotype is an independent risk factor for tuberculosis treatment failure in Indonesia', *J Infect Dis*, vol. 201, no. 4, pp. 553-557.

Pérez-Lago, L, Comas, I, Navarro, Y, González-Candelas, F, Herranz, M, Bouza, E & García-de-Viedma, D 2014, 'Whole-genome Sequencing Analysis of Intrapatient Microevolution in *Mycobacterium tuberculosis*: Potential Impact on the Inference of Tuberculosis Transmission', *J Infect Dis*, vol. 209, no. 1, pp. 98-108.

Periwal, V, Patowary, A, Vellarikkal, SK, Gupta, A, Singh, M, Mittal, A, Jeyapaul, S, Chauhan, RK, Singh, AV, Singh, PK, Garg, P, Katoch, VM, Katoch, K, Chauhan, DS, Sivasubbu, S & Scaria, V 2015, 'Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of *Mycobacterium tuberculosis* Pangenome', *PLoS One*, vol. 10, no. 4, doi: 10.1371/journal.pone.0122979.

Piccazzo, R, Paparo, F & Garlaschi, G 2014, 'Diagnostic Accuracy of Chest Radiography for the Diagnosis of Tuberculosis (TB) and Its Role in the Detection of Latent TB Infection: a Systematic Review', *J Rheumatol*, vol. 91 (supplement), pp. 32-40.

Pollett, S, Banner, P, O'Sullivan, MVN & Ralph, AP 2016, 'Epidemiology, Diagnosis and Management of Extra-Pulmonary Tuberculosis in a Low-

Prevalence Country: A Four Year Retrospective Study in an Australian Tertiary Infectious Diseases Unit', *PLoS One*, vol. 11, no. 3, doi: 10.1371/journal.pone.0149372.

Pouseele H & Supply P 2015, 'Accurate Whole-Genome Sequencing Based Epidemiological Surveillance of *Mycobacterium tuberculosis*', Methods in Microbiology (vol. 2): Oxford Academic Press.

Powell, KM, VanderEnde, DS, Holland, DP, Haddad, MB, Yarn, B, Yamin, AS, Mohamed, O, Sales, RF, DiMiceli, LE, Burns-Grant, G, Reaves EJ, Gardner TJ & Ray, SM 2017, 'Outbreak of Drug-Resistant *Mycobacterium tuberculosis* Among Homeless People in Atlanta, Georgia, 2008-2015', *Public Health Rep*, vol. 132, no. 2, pp. 231-240.

Raviglione, M 2006, 'XDR-TB: entering the post-antibiotic era?', *Int J Tuberc Lung Dis*, vol. 10, no. 11, pp. 1185-1187.

Raynaud, C, Guilhot, C, Rauzier, J, Bordat, Y, Pelicic, V, Manganelli, R., Smith, I, Gicquel, B & Jackson, M 2002, 'Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*', *Mol Microbiol*, vol. 45, no. 1, pp. 203-217.

Reed, MB, Pichler, VK, McIntosh, F, Mattia, A, Fallow, A, Masala, S, Domenech, P, Zwerling, A, Thibert, L, Menzies, D, Schwartzman, K, Behr, MA 2009, 'Major *Mycobacterium tuberculosis* Lineages Associate with Patient Country of Origin', *J Clin Microbiol*, vol. 47, no. 4, pp. 1119-1128.

Reichman, LB 1997, 'Tuberculosis elimination--what's to stop us?' *Int J Tuberc Lung Dis*, vol. 1, no.1, pp. 3-11.

Reuter, S, Harrison, TG, Köser, CU, Ellington, MJ, Smith, GP, Parkhill, J, Peacock, SJ, Bentley, SD & Török, ME 2013, 'A pilot study of rapid whole-genome sequencing for the investigation of a Legionella outbreak', B*MJ Open*, vol. 3, no. 1, doi:10.1136/bmjopen-2012-002175.

Riojas, MA, McGough, KJ, Rider-Riojas, CJ, Rastogi, N & Hazbón, MH 2018, 'Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of *Mycobacterium tuberculosis*', *Int J Syst Evol Microbiol,* vol. 68, no.1, pp. 324-332.

Ricks, PM, Cain, KP, Oeltmann, JE, Kammerer, JS & Moonan, PK 2011, 'Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the US, 2005–2009', *PLoS One*, vol. 6, no. 11, doi: 10.1371/journal.pone.0027405.

Robin, JD, Ludlow, AT, LaRanger, R, Wright, WE & Shay, JW 2016, 'Comparison of DNA Quantification Methods for Next Generation Sequencing', *Sci Rep*, vol. 6, doi:10.1038/srep24067.

Rodrigues, LC, Mangtani, P & Abubakar, I 2011, 'How does the level of BCG vaccine protection against tuberculosis fall over time?' *BMJ*, vol. 343, doi: 10.1136/bmj.d5974.

Rodwell, TC, VF, Douglas, J, Qian, L, Garfein, RS, Chawla, A, Torres J, Zadorozhny, V, Kim, MS, Hoshide, M, Catanzaro, D, Jackson, L, Lin, G, Desmond, E, Rodrigues, C, Eisenach, K, Victor, TC, Ismail N, Crudu, V, Gler, MT & Catanzaro, A 2014, 'Predicting Extensively Drug-Resistant

Mycobacterium tuberculosis Phenotypes with Genetic Mutations', *J Clin Microbiol*, vol. 52, no. 3, pp. 781-789.

Roe, M & Wood-Baker R 1999, 'Life over Death: Tasmanians and Tuberculosis', Tasmania: Tasmanian Historical Research Association.

Roetzer, A, Diel, R, Kohl, TA, Rückert, C, Nübel, U, Blom, J, Wirth, T, Jaenicke, S, Schuback, S, Rüsch-Gerdes, S, Supply, P, Kalinowski, J & Niemann, S 2013, 'Whole-genome Sequencing versus Traditional Genotyping for Investigation of a *Mycobacterium tuberculosis* Outbreak: A Longitudinal Molecular Epidemiological Study', *PLoS Med*, vol. 10, no. 2. doi: 10.1371/journal.pmed.1001387

Roetzer, A, Schuback, S, Diel, R, Gasau, F, Ubben, T, di Nauta, A, Richter, E, Rüsch-Gerdes, S & Niemann, S 2011, 'Evaluation of *Mycobacterium tuberculosis* Typing Methods in a 4-Year Study in Schleswig-Holstein, Northern Germany', *J Clin Microbiol*, vol. 49, no. 12, pp. 4173-4178.

Roy, A, Eisenhut, M, Harris, RJ, Rodrigues, LC, Sridhar, S, Habermann, S, Snell, L, Mangtani, P, Adetifa, I, Lalvani, A & Abubakar, I 2014, 'Effect of BCG vaccination against *Mycobacterium tuberculosis* infection in children: systematic review and meta-analysis', *BMJ*, vol. 349, doi:10.1136/bmj.g4643.

Rozenshtein, A, Hao, F, Starc, MT & Pearson, GDN 2015, 'Radiographic Appearance of Pulmonary Tuberculosis: Dogma Disproved', *AJR Am J Roentgenol*, vol. 204, no. 5, pp. 974-978.

Ryan, F 1994, ‘The Forgotten Plague: how the battle against tuberculosis was won-and lost’, Boston, MA: Back Bay Books.

Saelens, JW, Lau-Bonilla, D, Moller, A, Medina, N, Guzmán, B, Calderón, M, Herrera, R, Sisk, DM, Xet-Mull, AM, Stout, JE, Arathoon, E, Samayoa, B & Tobin, DM 2015, ‘Whole-genome sequencing identifies circulating Beijing-lineage *Mycobacterium tuberculosis* strains in Guatemala and an associated urban outbreak’, *Tuberculosis (Edinb)*, vol. 95, no. 6, pp. 810-816.

Sanchez-Padilla, E, Merker, M, Beckert, P, Jochims, F, Dlamini, T, Kahn, P, Bonnet M & Niemann, S 2015, ‘Detection of drug-resistant tuberculosis by Xpert MTB/RIF in Swaziland’, *N Engl J Med*, vol. 372, no. 12, pp. 1181-1182.

Sandgren A, Schepisi, MS, Sotgiu, G, Huitric, E, Migliori, GB, Manissero, D, van der Werf, MJ & Girardi E 2014, ‘Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: a systematic review’, *Eur Respir J*, vol. 43, no. 4, pp. 1159-1171.

Savine, E, Warren, RM, van der Spuy, GD, Beyers, N, van Helden, PD, Locht, C & Supply, P 2002, ‘Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis*’, *J Clin Microbiol*, vol. 40, no. 12, pp. 4561-4566.

Schleusener, V, Koser, CU, Beckert, P, Niemann, S & Feuerriegel, S 2017, ‘*Mycobacterium tuberculosis* resistance prediction and lineage classification

from genome sequencing: comparison of automated analysis tools', *Sci Rep*, vol. 7, doi:10.1038/srep46327.

Schork, NJ, Fallin, D & Lanchbury, JS 2000, 'Single nucleotide polymorphisms and the future of genetic epidemiology', *Clin Genet*, vol. 58, no. 4, pp. 250-264.

Schrager, L, Olesen, O, Vekemans, J, Lewinsohn, D, Shea, J, Hanekom, W & Drager, N 2018, 'Global Report on Tuberculosis Vaccines', [Electronic edition] Retrieved from: https://www.tbvi.eu/wp-content/uploads/2018/02/Summary-SWRTV_Finalproof.pdf

Schreuder, LJ, Carroll, P, Muwanguzi-Karugaba, J, Kokoczka, R, Brown, AC & Parish, T 2015, '*Mycobacterium tuberculosis* H37Rv has a single nucleotide polymorphism in PhoR which affects cell wall hydrophobicity and gene expression', *Microbiology*, vol. 161, no. 4, pp. 765-773.

Schwartzman, K & Menzies, D 2000, 'How long are TB patients infectious?', *CMAJ*, vol. 163. no. 2, pp. 157-158.

Schwarze, K, Buchanan, J, Taylor, JC & Wordsworth, S 2018, 'Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature', *Genet Med*, vol. 20, pp. 1122-1130.

Sekizuka, T, Yamashita, A, Murase, Y, Iwamoto, T, Mitarai, S, Kato, S & Kuroda, M 2015, 'TGS-TB: Total Genotyping Solution for *Mycobacterium tuberculosis* Using Short-Read Whole-Genome Sequencing', *PLoS One*, vol. 10, no. 11, doi: 10.1371/journal.pone.0142951.

Sexton, K, Petera, S & Pandey, S 2008, 'Five years of molecular typing of *Mycobacterium tuberculosis* isolates in New Zealand, 2003 to 2007', [Electronic edition] Retrieved from: https://surv.esr.cri.nz/PDF_surveillance/TbMolTyp/TBMolecTypReport.pdf?m=1221688560

Sezonov, G, Joseleau-Petit, D & D'Ari, R 2007, 'Escherichia coli Physiology in Luria-Bertani Broth', *J Bacteriol*, vol. 189, no. 23, pp. 8746-8749.

Simons, S, van der Laan, T, Mulder, A, van Ingen, J, Rigouts, L, Dekhuijzen, P, Boeree MJ & van Soolingen, D 2014, 'Rapid diagnosis of pyrazinamide-resistant multidrug-resistant tuberculosis using a molecular-based diagnostic algorithm', *Clin Microbiol Infect*, vol. 20, no. 10, pp. 1015-1020.

Skoura, E, Zumla, A & Bomanji, J 2015, 'Imaging in tuberculosis', *Int J Infect Dis*, vol. 32, pp. 87-93.

Small, PM, Shafer, RW, Hopewell, PC, Singh, SP, Murphy, MJ, Desmond, E, Sireea, MF & Schoolnik, GK 1993, 'Exogenous Reinfection with Multidrug-Resistant *Mycobacterium tuberculosis* in Patients with Advanced HIV Infection', *N Engl J Med*, vol. 328, no. 16, pp. 1137-1144.

Sobkowiak, B, Glynn, JR, Houben, RMGJ, Mallard, K, Phelan, JE, Guerra-Assunção, JA, Banda, L, Mzembe, T, Viveiros, M, McNerney, R, Parkhill, J, Crampin, AC & Clark, TG 2018, 'Identifying mixed *Mycobacterium tuberculosis* infections from whole-genome sequence data', *BMC Genomics*, vol. 19, no. 1, doi:10.1186/s12864-018-4988-z.

Solans, L, Aguiló, N, Samper, S, Pawlik, A, Frigui, W, Martín, C, Brosch R & Gonzalo-Asensio, J 2014, 'A Specific Polymorphism in *Mycobacterium tuberculosis* H37Rv Causes Differential ESAT-6 Expression and Identifies WhiB6 as a Novel ESX-1 Component', *Infect Immun*, vol. 82, no. 8, pp. 3446-3456.

Sreevatsan, S, Stockbauer, KE, Pan, X, Kreiswirth, BN., Moghazeh, SL, Jacobs, WR Jr, Telenti A & Musser, JM 1997, 'Ethambutol resistance in *Mycobacterium tuberculosis*: critical role of embB mutations', *Antimicrob Agents Chemother*, vol. 41, no. 8, pp. 1677-1681.

Starks, AM, Gumusboga, A, Plikaytis, BB, Shinnick, TM & Posey, JE 2009, 'Mutations at embB codon 306 are an important molecular indicator of ethambutol resistance in *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 53, no. 3, pp. 1061-1066.

Struelens, MJ 1998, 'Molecular epidemiologic typing systems of bacterial pathogens: current issues and perspectives', *Mem Inst Oswaldo Cruz*, vol. 93, no. 5, pp. 581-585.

Stucki, D, Brites, D, Jeljeli, L, Coscolla, M, Liu, Q, Trauner, A, Fenner, L, Rutaihwa, L, Borrell, S, Luo, T, Gao, Q, Kato-Maeda, M, Ballif, M, Egger, M, Macedo, R, Mardassi, H, Moreno, M, Tudo Vilanova, G, Fyfe, J, Globan, M, Thomas, J, Jamieson, F, Guthrie, JL, Asante-Poku, A, Yeboah-Manu, D, Wampande, E, Ssengooba, W, Joloba, M, Henry Boom, W, Basu, I, Bower, J, Saraiva, M, Vaconcellos, SEG, Suffys, P, Koch, A, Wilkinson, R, Gail-Bekker, L, Malla, B, Ley, SD, Beck, HP, de Jong, BC, Toit, K, Sanchez-Padilla, E, Bonnet, M, Gil-Brusola, A, Frank, M, Penlap Beng, VN, Eisenach,

K, Alani, I, Wangui Ndung'u, P, Revathi, G, Gehre, F, Akter, S, Ntoumi, F, Stewart-Isherwood, L, Ntinginya, NE, Rachow, A, Hoelscher, M, Cirillo, DM, Skenders, G, Hoffner, S, Bakonyte, D, Stakenas, P, Diel, R, Crudu, V, Moldovan, O, Al-Hajoj, S, Otero, L, Barletta, F, Jane Carter, E, Diero, L, Supply, P, Comas, I, Niemann, S & Gagneux, S 2016, '*Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages', *Nat Genet*, vol. 48, no. 12, pp. 1535-1543.

Supply, P, Allix, C, Lesjean, S, Cardoso-Oelemann, M, Rusch-Gerdes, S, Willery, E, Savine, E, de Haas, P, van Deutekom, H, Roring, S, Bifani, P, Kurepina, N, Kreiswirth, B, Sola, C, Rastogi, N, Vatin V, Gutierrez, MC, Fauville, M, Niemann, S, Skuce, R, Kremer, K, Locht, C & van Soolingen, D 2006, 'Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*', *J Clin Microbiol*, vol. 44, no. 12, pp. 4498-4510.

Swaminathan, S & Rekha, B 2010, 'Pediatric tuberculosis: global overview and challenges', *Clin Infect Dis*, vol. 50(supplement 3), pp. S184-194.

Takiff, HE, Salazar, L, Guerrero, C, Philipp, W, Huang, WM, Kreiswirth, B, Cole, ST, Jacobs, WR Jr & Telenti, A 1994, 'Cloning and nucleotide sequence of *Mycobacterium tuberculosis* gyrA and gyrB genes and detection of quinolone resistance mutations', *Antimicrob Agents Chemother*, vol. 38, no. 4, pp. 773-780.

Tatusov, RL, Galperin, MY, Natale, DA & Koonin, EV 2000, 'The COG database: a tool for genome-scale analysis of protein functions and evolution', *Nucleic Acids Res*, vol. 28, no. 1, pp. 33-36.

Telenti, A, Imboden, P, Marchesi, F, Lowrie, D, Cole, S, Colston, MJ, Schopfer K & Bodmer, T 1993, 'Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*', *Lancet*, vol. 341, no. 8846, pp. 647-650.

Tewolde, E, Atnafu A, Kebede, T, Tedla, Y, Assefa, G, Bedane, A & Enquselassie F 2015, 'Evaluating the validity and reliability of chest radiography in the diagnosis of tuberculosis among smear negative pulmonary tuberculosis patients', *Ethiop Med J*, vol. 53, no. 2, pp. 83-89.

Thrupp, L, Bradley, S, Smith, P, Simor, A, Gantz, N, Crossley, K, Loeb, M, Strausbaugh, L, Nicolle L & SHEA Long-Term-Care Committee. 2004, 'Tuberculosis Prevention and Control in Long-Term–Care Facilities for Older Adults', Infect Control Hosp Epidemiol, vol. 25, no. 12, pp. 1097-1108.

Toms, C, Stapledon, R, Coulter, C & Douglas, P 2017, 'Tuberculosis notifications in Australia, 2014', *Commun Dis Intell Q Rep*, vol. 41, no. 3, pp. e247-e263.

Török, ME, Reuter, S, Bryant, J, Köser, CU, Stinchcombe, SV, Nazareth, B, Ellington, MJ, Bentley, SD, Smith, GP, Parkhill, J & Peacock, SJ 2013, 'Rapid Whole-Genome Sequencing for Investigation of a Suspected Tuberculosis Outbreak', *J Clin Microbiol*, vol. 51, no. 2, pp. 611-614.

Trunz, BB, Fine, P & Dye, C 2006, 'Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness', Lancet, vol. 367, no. 9517, pp. 1173-1180.

Tsenova, L, Ellison, E, Harbacheuski, R, Moreira, AL, Kurepina, N, Reed, MB, Mathema, B, Barry, CE 3rd & Kaplan, G 2005, 'Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli', *J Infect Dis*, vol. 192, no. 1, pp. 98-106.

Ulrichs, T, & Kaufmann, SH, 2006, 'New insights into the function of granulomas in human tuberculosis', *J Pathol*, vol. 208, no. 2, pp. 261-269.

Valway, SE, Sanchez, MPC, Shinnick, TF, Orme, I, Agerton, T, Hoy, D, Jones, JS, Westmoreland, H & Onorato, IM 1998, 'An Outbreak Involving Extensive Transmission of a Virulent Strain of *Mycobacterium tuberculosis*', *N Engl J Med*, vol. 338, no.10, pp. 633-639.

van Duin, JM, Pijnenburg, JE, van Rijswoud, CM, de Haas, PE, Hendriks, WD & van Soolingen, D 1998, 'Investigation of cross contamination in a *Mycobacterium tuberculosis* laboratory using IS6110 DNA fingerprinting', *Int J Tuberc Lung Dis*, vol. 2, no. 5, pp. 425-429.

van Embden, JD, Cave, MD, Crawford, JT, Dale, JW, Eisenach, KD, Gicquel, B, Martin, C, McAdam R & Shinnick, TM 1993, 'Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology', *J Clin Microbiol*, vol. 31, no. 2, pp. 406-409.

van Leth, F, van der Werf, MJ & Borgdorff, MW 2008, 'Prevalence of tuberculous infection and incidence of tuberculosis: a re-assessment of the Styblo rule', *Bull World Health Organ*, vol. 86, no. 1, pp. 20-26.

van Soolingen, D, Borgdorff, MW, de Haas, PE, Sebek, MM, Veen, J, Dessens, M, Kremer K & van Embden, JD 1999, 'Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997', *J Infect Dis*, vol. 180, no. 3, pp. 726-736.

Veen, J 1995, 'Drug resistant tuberculosis: back to sanatoria, surgery and cod-liver oil?', *Eur Respir J*, vol. 8, no. 7, pp.1073-1075.

Votintseva, AA, Bradley, P, Pankhurst, L, del Ojo Elias, C, Loose, M, Nilgiriwala, K, Chatterjee, A, Smith, EG, Sanderson, N, Walker, TM, Morgan, MR, Wyllie, DH, Walker, AS, Peto, TEA, Crook, DW & Iqbal, Z 2017, 'Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples', *J Clin Microbiol*, vol. 55, no. 5, pp. 1285-1298.

Walker, TM, Ip, CL, Harrell, RH, Evans, JT, Kapatai, G, Dedicoat, MJ, Eyre, DW, Wilson, DJ, Hawkey, PM, Crook, DW, Parkhill, J, Harris, D, Walker, AS, Bowden, R, Monk, P, Smith, EG & Peto, TE 2013, 'Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study', *Lancet Infect Dis*, vol. 13, no. 12, pp. 137-146.

Walker, TM, Kohl, TA, Omar, SV, Hedge, J, Del Ojo Elias, C, Bradley, P, Iqbal, Z, Feuerriegel, S, Niehaus, KE, Wilson, DJ, Clifton, DA, Kapatai, G, Ip, CL, Bowden, R, Drobniewski, FA, Allix-Béguec, C, Gaudin, C, Parkhill, J, Diel, R, Supply, P, Crook, DW, Smith, EG, Walker, AS, Ismail, N, Niemann, S, Peto, TEA & Modernizing Medical Microbiology Informatics Group 2015, 'Whole-genome sequencing for prediction of *Mycobacterium*

tuberculosis drug susceptibility and resistance: a retrospective cohort study', *Lancet Infect Dis*, vol. 15, no. 10, pp.1193-1202.

Walters, SB, Dubnau, E, Kolesnikova, I, Laval, F, Daffe, M & Smith, I 2006, 'The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis', *Mol Microbiol*, vol. 60, no. 2, pp. 312-330.

Warren, RM, van der Spuy, GD, Richardson, M, Beyers, N, Booysen, C, Behr, MA & van Helden, PD 2002, 'Evolution of the IS6110-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*', *J Clin Microbiol*, 40, no. 4, pp. 1277-1282.

Warren, RM, van der Spuy, GD, Richardson, M, Beyers, N, Borgdorff, M. W, Behr, MA & van Helden, PD 2002, 'Calculation of the stability of the IS6110 banding pattern in patients with persistent *Mycobacterium tuberculosis* disease', *J Clin Microbiol*, vol. 40, no. 5, pp. 1705-1708.

Waziri, NE, Cadmus, S, Nguku, P, Fawole, O, Owolodun, OA, Waziri, H, Ibrahim, L, Biya, O, Gidado, S, Badung, S, Kumbish, P & Nsubuga, P 2014, 'Factors associated with tuberculosis among patients attending a treatment centre in Zaria, North-west Nigeria, 2010', *Pan Afr Med J*, vol. 18(supplement 1), doi: 10.11694/pamj.supp.2014.18.1.4189

Wells, WA, Boehme, CC, Cobelens, FG, Daniels, C, Dowdy, D, Gardiner, E, Gheuens, J, Kim, P, Kimerling, ME, Kreiswirth, B, Lienhardt, C, Mdluli, K, Pai, M, Perkins, MD, Peter, T, Zignol, M, Zuml,a A & Schito, M 2013, 'Alignment of new tuberculosis drug regimens and drug susceptibility

testing: a framework for action', *Lancet Infect Dis*, vol. 13, no. 5, pp. 449-458.

Weniger, T, Krawczyk, J, Supply, P, Niemann, S & Harmsen, D 2010, 'MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria', *Nucleic Acids Res*, vol. 38 (web server issue, W326-W3310, doi:10.1093/nar/gkq351.

Wetterstrand, KA 2018, 'DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program', [Electronic edition] Retrieved from: www.genome.gov/sequencingcostsdata

White, RG & Houben, RM 2014, 'Towards elimination in industrialised countries: expanding diagnosis and treatment of LTBI among immigrants', *Int J Tuberc Lung Dis*, vol. 18, no. 4, pp. 380-380.

Wieland, ML, Weis, JA, Yawn, BP, Sullivan, SM, Millington, KL, Smith, CM, Bertram, S, Nigon, JA & Sia, IG 2012, 'Perceptions of tuberculosis among immigrants and refugees at an adult education center: a community-based participatory research approach', *J Immigr Minor Health*, vol. 14, no. 1, pp. 14-22.

Wielgoss, S, Barrick, JE, Tenaillon, O, Cruveiller, S, Chane-Woon-Ming, B, Médigue, C, Lenski RE & Schneider, D 2011, 'Mutation Rate Inferred from Synonymous Substitutions in a Long-Term Evolution Experiment with Escherichia coli', *G3 (Bethesda)*, vol.1, no. 3, pp. 183-186.

Williams, M, Mizrahi, V & Kana, BD 2014, 'Molybdenum cofactor: a key component of *Mycobacterium tuberculosis* pathogenesis?' *Crit Rev Microbiol*, vol. 40, no. 1, pp. 18-29.

Williams, MJ, Kana, BD & Mizrahi, V 2011, 'Functional Analysis of Molybdopterin Biosynthesis in Mycobacteria Identifies a Fused Molybdopterin Synthase in *Mycobacterium tuberculosis*', *J Bacteriol*, vol. 193, no. 1, pp. 98-106.

Williams, MJ, Shanley, CA, Zilavy, A, Peixoto, B, Manca, C, Kaplan, G, Orme, IM, Mizrahi V & Kana, BD 2015, 'bis-Molybdopterin guanine dinucleotide is required for persistence of *Mycobacterium tuberculosis* in guinea pigs', *Infect Immun*, vol. 83, no. 2, pp. 544-550.

Wirth, T, Hildebrand, F, Allix-Béguec, C, Wölbeling, F, Kubica, T, Kremer, K, van Soolingen, D, Rüsch-Gerdes, S, Locht, C, Brisse, S, Meyer, A, Supply, P & Niemann, S 2008, 'Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex', *PLoS Pathog*, vol. 4, no. 9, doi:10.1371/journal.ppat.1000160

Witney, AA, Gould, KA, Arnold, A, Coleman, D, Delgado, R, Dhillon, J, Pond, MJ, Pope, CF, Planche, TD, Stoker, NG, Cosgrove, CA, Butcher, PD, Harrison, TS & Hinds, J 2015, 'Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases', *J Clin Microbiol*, vol. 53, no. 5, pp. 1473-1483.

Wollenberg, KR, Desjardins, CA, Zalutskaya, A, Slodovnikova, V, Oler, AJ, Quiñones, M, Abeel, T, Chapman, SB, Tartakovsky, M, Gabrielian, A, Hoffner, S, Skrahin, A, Birren, BW, Rosenthal, A, Skrahina, A & Earl, AM

2017, ‘Whole-Genome Sequencing of *Mycobacterium tuberculosis* Provides Insight into the Evolution and Genetic Composition of Drug-Resistant Tuberculosis in Belarus’, *J Clin Microbiol*, vol. 55, no. 2, pp. 457-469.

Workicho, A, Kassahun, W & Alemseged, F 2017, ‘Risk factors for multidrug-resistant tuberculosis among tuberculosis patients: a case-control study’, *Infect Drug Resist*, vol. 10, pp. 91-96.

World Health Organization 2012, ‘Chest Radiography in Tuberculosis Detection’, [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/252424/9789241511506-eng.pdf?sequence=1

World Health Organization 2013, ‘Systematic screening for active tuberculosis Principles and recommendations’, [Electronic edition] Retrieved from: https://www.who.int/tb/publications/Final_TB_Screening_guidelines.pdf

World Health Organization 2013a, ‘Recommendations to assure the quality, safety and efficacy of BCG vaccines’, [Electronic edition] Retrieved from: http://www.who.int/biologicals/BCG_DB_HK_23_April_2012.pdf

World Health Organization 2013b, ‘The use of bedaquiline in the treatment of multidrug-resistant tuberculosis Interim policy guidance’, [Electronic edition] Retrieved from: http://www.wpro.who.int/philippines/topics/tuberculosis/Bedaquiline_Use_interim_policy.pdf

World Health Organization 2014, ‘Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis’, [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/130918/9789241548809_eng.pdf?sequence=1

World Health Organization 2014a, ‘Definitions and reporting framework for tuberculosis-2013 revision,’ [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/79199/9789241505345_eng.pdf?sequence=1

World Health Organization 2014b, ‘Guidance for national tuberculosis programmes on the management of tuberculosis in children’, [Electronic edition] Retrieved from: http://apps.who.int/medicinedocs/documents/s21535en/s21535en.pdf

World Health Organization 2014c, ‘The use of delamanid in the treatment of multidrug-resistant tuberculosis’, [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/137334/WHO_HTM_TB_2014.23_eng.pdf?sequence=1

World Health Organization 2015, ‘Frequently asked questions about the implementation of the new WHO recommendation on the use of the shorter MDR-TB regimen under programmatic conditions’, [Electronic edition] Retrieved from: http://www.who.int/tb/areas-of-work/drug-resistant-tb/treatment/FAQshorter_MDR_regimen.pdf?ua=1

World Health Organization 2015a, ‘Guidelines on the management of latent tuberculosis infection’, [Electronic edition] Retrieved from:

http://apps.who.int/iris/bitstream/handle/10665/136471/9789241548908_eng.pdf?sequence=1

World Health Organization 2015b, ‘Use of high burden country lists for TB by WHO in the post-2015 era’, [Electronic edition] Retrieved from: http://www.who.int/tb/publications/global_report/high_tb_burdencountrylists2016-2020summary.pdf

World Health Organization 2016, ‘Global Tuberculosis Report 2016’, [Electronic edition] Retrieved from: http://apps.who.int/medicinedocs/documents/s23098en/s23098en.pdf

World Health Organization 2016a, ‘WHO treatment guidelines for drug-resistant tuberculosis’, [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/250125/9789241549639-eng.pdf?sequence=1

World Health Organization 2017, ‘Global Tuberculosis Report 2017’, [Electronic edition] Retrieved from: http://www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf

World Health Organization 2017a, ‘Guidelines for treatment of drug-susceptible tuberculosis and patient care’, [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/255052/9789241550000-eng.pdf?sequence=1

World Health Organization 2017b, ‘Report on BCG vaccine use for protection against mycobacterial infections including tuberculosis, leprosy,

and other nontuberculous mycobacteria (NTM) infections', [Electronic edition] Retrieved from: http://www.who.int/immunization/sage/meetings/2017/october/1_BCG_report_revised_version_online.pdf

World Health Organization 2017c, 'Tuberculosis', [Electronic edition] Retrieved from: http://www.who.int/tb/publications/factsheet_global.pdf?ua=1

World Health Organization 2018, 'Australia', [Electronic edition] Retrieved from: https://extranet.who.int/sree/Reports?op=Replet&name=%2FWHO_HQ_Reports%2FG2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=AU&LAN=EN&outtype=html

World Health Organization 2018a, 'BCG vaccine', [Electronic edition] Retrieved from: http://www.who.int/biologicals/areas/vaccines/bcg/en/

World Health Organization 2018b, 'Drug-resistant TB: Totally drug-resistant TB FAQ', [Electronic edition] Retrieved from: http://www.who.int/tb/areas-of-work/drug-resistant-tb/totally-drug-resistant-tb-faq/en/

World Health Organization 2018c, 'Global Health Observatory (GHO) data', [Electronic edition] Retrieved from: http://www.who.int/gho/hiv/epidemic_status/deaths_text/en/

World Health Organization 2018d, 'Global Tuberculosis Report 2018', [Electronic edition] Retrieved from:

http://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf?ua=1

World Health Organization 2018e, ‘Latent tuberculosis infection Updated and consolidated guidelines for programmatic management’, [Electronic edition] Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/260233/9789241550239-eng.pdf?sequence=1

World Health Organization 2018f, ‘Tuberculosis’, [Electronic edition] Retrieved from: http://www.who.int/tb/areas-of-work/drug-resistant-tb/xdr-tb-faq/en/

World Health Organization 2018g, ‘Viet Nam’, [Electronic edition] Retrieved from: https://extranet.who.int/sree/Reports?op=Replet&name=%2FWHO_HQ_Reports%2FG2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=VN&LAN=EN&outtype=html

World Health Organization 2018h, ‘New Zealand’, [Electronic edition] Retrieved from: https://extranet.who.int/sree/Reports?op=Replet&name=%2FWHO_HQ_Reports%2FG2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=NZ&LAN=EN&outtype=html

World Health Organization 2018i, ‘What is multidrug-resistant tuberculosis (MDR-TB) and how do we control it?’, [Electronic edition] Retrieved from: http://www.who.int/features/qa/79/en/

Yang, C, Luo, T, Sun, G, Qiao, K, Sun, G, DeRiemer, K, Mei, J & Gao, Q 2012, '*Mycobacterium tuberculosis* Beijing Strains Favor Transmission but Not Drug Resistance in China', *Clin Infect Dis*, vol. 55, no. 9, pp. 1179-1187.

Yen. S, Brower, JE, Freeman, JT, Basu, I, O'Toole, RF 2013, 'Phylogenetic lineages of tuberculosis isolates in New Zealand and their association with patient demographics', *Int J Tuberc Lung Dis*, vol. 17, no. 7, pp. 892-897.

Zhang, M, Gong, J, Lin, Y & Barnes, PF 1998, 'Growth of virulent and avirulent *Mycobacterium tuberculosis* strains in human macrophages', *Infect Immun*, vol. 66, no. 2, pp. 794-799.

Zhang, SL, Zhao, JW, Sun, ZQ, Yang, EZ, Yan, JH, Zhao, Q, Zhang, GL, Zhang, HM, Qi, YM, Wang, HH & Sun, QW 2009, 'Development and evaluation of a novel multiple-antigen ELISA for serodiagnosis of tuberculosis', *Tuberculosis (Edinb)*, vol. 89, no. 4, pp. 278-284.

Zhao, F, Wang, XD, Erber, LN, Luo, M, Guo, AZ., Yang, SS, Gu J, Turman, BJ, Gao, YR, Li, DF, Cui, ZQ, Zhang, ZP, Bi, LJ, Baughn, AD, Zhang, XE & Deng, JY 2014, 'Binding pocket alterations in dihydrofolate synthase confer resistance to para-aminosalicylic acid in clinical isolates of *Mycobacterium tuberculosis*', *Antimicrob Agents Chemother*, vol. 58, no. 3, pp.1479-1487.

Supplement 1

Step by step guide to analyse the Fastq files using Geneious

1. Download Geneious from https://www.geneious.com/

2. Go to File | Import | From File. Import raw-read files (Sample_xx_R1.fastq.gz and Sample_xx_R2.fastq.gz) into Geneious

3. Download the Reference Genome from the NCBI database. For example, *M. tuberculosis* NC_000962.3.

a. In the Left panel | Go to NCBI | Nucleotide.

b. Enter NC_000962.3 | Click Search.

c. Once the genome has been found, click Download Full Sequence(s).

d. Download the NC_000962.3 reference genome (The icon changes to a green circular genome when completed).

e. Drag and drop the NC_000962.3 reference genome into the working folder.

4. Mapping the isolate sequence to the reference genome

a. Hold CTRL and select the both R1 and R2 raw read files (imported), and the reference genome (NC_000962.3) (downloaded).

b. Click Align| Assemble| Map to Reference.

c. Check the settings*

i. Reference Sequence = NC_000962.3

ii. Mapper = Bowtie2 – fast and accurate read mapper

iii. Trim Before Mapping = Do not trim

iv. Results: Select all options

v. Results | Save consensus sequences | Options

vi. Threshold = Highest Quality

vii. Threshold for sequences without quality = 95%

viii. If no coverage call = '–'

d. When mapping to reference is complete, a new folder will be created containing four files:

i. Assembly Report

ii. Consensus

iii. Contig

iv. Unused Reads

e. A contig is contiguous linear stretch of DNA or RNA consensus sequence constructed from a number of smaller, partially overlapping, sequence fragments (reads). To view Contigs, click Contig file | Slide the zoom bar to view the sequence reads. The statistics about the mapping can be viewed by clicking % tab in view statistics.

*This setting may vary depending on the objective of analyses and quality of Fastq reads.

Supplement 2

Publications

Taylor & Francis
Taylor & Francis Group

INFECTIOUS DISEASES,
2017; VOL. 49,
NO. 9, 680–688

https://doi.org/10.1080/23744235.2017.1330553

ORIGINAL ARTICLE

Check for updates

Differential carriage of virulence-associated loci in the New Zealand Rangipo outbreak strain of *Mycobacterium tuberculosis*

Sanjay S. Gautam[a], Micheál Mac Aogáin[b], James E. Bower[c], Indira Basu[c] and Ronan F. O'Toole[a]

[a]School of Medicine, University of Tasmania, Hobart, Australia; [b]Department of Clinical Microbiology, Trinity Translational Medicine Institute, School of Medicine, Trinity College Dublin, St. James's Hospital, Dublin, Ireland; [c]LabPLUS, Auckland City Hospital, Auckland, New Zealand

ABSTRACT

Background: The Rangipo strain of *Mycobacterium tuberculosis* achieved notoriety in New Zealand due to its role in several tuberculosis (TB) outbreaks. Why this strain should be the source of relatively large clusters of the disease is unknown. In this work, we performed an in-depth analysis of the genome of the Rangipo strain to determine whether it offers clues to understanding its prevalence.

Methods: Next-generation sequencing was performed on nine isolates which matched the Rangipo genotypic profile. Sequence reads were assembled against the H37Rv reference genome and single-locus variants identified. Unmapped reads were compared against the genome sequences of other *M. tuberculosis* strains, in particular CDC1551, Haarlem and Erdman.

Results: Across the nine Rangipo strains, a total of 727 single-locus variants were identified with respect to H37Rv, of which 700 were common to all Rangipo strains sequenced. Within the common variants, 386 were non-synonymous, with 12 occurring in genes associated with *M. tuberculosis* virulence. Next-generation and Sanger sequencing determined the presence of three genes in the Rangipo isolates, which are absent in H37Rv, but which have been reported to be important for the pathogenicity of *M. tuberculosis*. The differentially encoded Rangipo genes consisted of transcriptional regulator EmbR2, and molybdopterin cofactor biosynthesis proteins A and B. The Rangipo strain also harbours an extended DNA helicase and an additional adenylate cyclase.

Conclusions: Our study provides new insights into the genomic content of the New Zealand Rangipo strain of *M. tuberculosis* and highlights the presence of additional virulence-related loci not found in H37Rv.

KEYWORDS
Mycobacterium tuberculosis
CDC1551
EmbR2
whole-genome sequencing

ARTICLE HISTORY
Received 27 December 2016
Revised 20 April 2017
Accepted 10 May 2017

CONTACT
Ronan F. O'Toole
ronan.otoole@utas.edu.au
School of Medicine, Faculty of Health,
University of Tasmania, Medical Science 1, 17
Liverpool Street, Hobart, TAS 7000, Australia

Supplemental data for this article can be accessed here.

Introduction

Tuberculosis (TB) is the leading cause of mortality due to respiratory infection worldwide, killing ~1.5 million people in 2014 alone [1]. New Zealand has a low notification rate of TB with 6.7 cases per 100,000 population (302 cases) in 2014 [2]; however, the distribution of the disease is not uniform across the population with some districts and at-risk groups exhibiting disparate TB rates. For example, the notification rate of TB in the Asian ethnic group was 56.8 times higher (34.1 per 100,000) than in the European or other ethnic group (0.6 per 100,000) in 2014 [2]. TB rates have also been consistently higher in Māori and Pacific peoples compared to New Zealanders of European descent [2,3]. Furthermore, intermittent outbreaks of TB can significantly impact the national rate from year to year.

In 2001, an analysis of an outbreak of TB in New Zealand which occurred between November 1996 and May 2000 was reported [4]. Forty-three of the 61 TB cases were confirmed as belonging to the outbreak based on IS6110-based restriction fragment length polymorphism (RFLP) typing of the isolates, with the remaining 18 cases determined by epidemiological contact tracing. One of the patients had previously served a prison sentence in the Tongariro/Rangipo Prison in 1998, and as a result, the strain subsequently became referred to as the Rangipo strain [5]. A subsequent outbreak in 2002 in the Hawke's Bay involving 19 active cases of Rangipo TB was associated with a high rate of transmission as determined by the presentation of TB disease or latent TB infection in 16.4% and 20.0%, respectively, of household or other close contacts [5]. A molecular typing study of *M. tuberculosis* complex isolates from 2003 to 2007 found that the Rangipo strain was associated with the highest number of district health boards (DHBs) in New Zealand (13 out of 18) compared to other clusters which were related to a median of 2 DHBs [6].

The reasons underlying the high infectivity and prevalence of the Rangipo strain are not known. Prison incarceration was a factor common to a number of cases [4]. The role of the prison environment in the epidemiology of TB has been well documented with active TB occurring at higher levels than among the general population [7]. This has been attributed to the fact that a disproportionate number of prisoners may exhibit TB risk factors including drug or alcohol misuse, homelessness or low socioeconomic status [8]. In addition, the location of prisoners is normally based on crime rather than public health considerations which can contribute to overcrowding, delayed diagnosis or inadequate treatment [7]. It is believed that prisons can act as reservoirs for TB transmission in the wider community [9].

It has been postulated that the Rangipo strain may be more infectious than other strains in circulation [5]; however, there are currently little published experimental data available in regard to this. We have previously shown that the Rangipo strain belongs to the Euro-American Lineage 4 of *Mycobacterium tuberculosis* complex based on earlier large sequence and single nucleotide polymorphic analyses [10]. In this work, we examined isolates of the Rangipo strain using whole-genome sequencing to shed light on potential differences in the encoded pathogenicity of this strain. In addition to single-locus variations, we discuss a series of virulence genes which distinguish the Rangipo strain from H37Rv and other reference *M. tuberculosis* strains.

Materials and methods

Bacterial strains

Mycobacterial Interspersed Repetitive Unit (MIRU) 24-loci typing was performed as previously described [11] on clinical isolates of *Mycobacterium tuberculosis* complex at LabPLUS, Auckland City Hospital. Isolates were defined as belonging to a cluster where they contained the same copy number at all loci. In this work, eight isolates with the identical 24-loci MIRU profile (233325153324 341444223362) of the Rangipo strain [5], and an additional ninth isolate, strain 356, which differed by one MIRU locus, were selected for whole-genome sequencing. The isolates analysed were from across multiple geographical locations and collected over a two-year period. The isolates were grown in Mycobacteria Growth Indicator Tube (MGIT) media and the cells lysed using glass beads for subsequent genomic DNA (gDNA) extraction.

Next-generation sequencing

gDNA of the *M. tuberculosis* isolates were purified, RNase-treated and quantified. The gDNA was tagged and amplified using a Nextera® XT DNA Library Preparation Kit, San Diego, CA and Nextera® XT Index Kit, San Diego, CA. The libraries generated were cleaned using Agencourt AMPure XP beads, normalized and then pooled. The concentration of the pooled library was determined by qPCR using a KAPA Library Quantification Kit, Wilmington, MA. 15 pM of the pooled library was loaded into the cartridge of a MiSeq Reagent Kit v2

which was run on an Illumina® MiSeq, San Diego, CA instrument and the generated FASTQ sequence files were collected.

Whole-genome sequence analysis

The FASTQ files of the nine Rangipo strains of *M. tuberculosis* were imported into the Geneious R9.0 software suite [12]. Paired-end sequence reads were trimmed (error probability limit of 0.05) and mapped (random multiple base matches) to the publicly available annotated genome of *M. tuberculosis* reference strain H37Rv (accession number NC_000962.3) [11] in the first instance. The maximum variant *p*-value was set at 10^{-6} when exceeding 65% bias. Single-locus variations (SLVs) were called at a minimum variant frequency of 95% and a minimum mean genome coverage of 20 and annotated as previously described [13]. Described repetitive regions in the H37Rv genome were masked to mitigate spurious variant calling [14]. A list of Rangipo common non-synonymous SLVs was created and classified based on the clusters of orthologous groups (COGs). Information on the known role of these genes with respect to mycobacterial virulence was obtained from the published literature.

Identification of differentially encoded genes in the Rangipo strain

To identify differentially encoded genes that are present in Rangipo, *de novo* assembly was performed on unmapped FASTQ reads from each of the Rangipo strains that did not map to the reference genome H37Rv using a trim error probability limit of 0.05. A consensus sequence was generated for each of the *de novo* assemblies and open reading frames (ORFs), with a minimum length of 150 nucleotides, standard genetic code and start codons ATG, TTG and CTG. Interior ORFs were included based on the assumption that start and/or stop codons could be located outside of the minimum ORF length. The predicted protein products of the ORFs were compared against other *M. tuberculosis* strains in the database of non-redundant protein sequences by BLASTP analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [15] to identify Rangipo strain orthologues of genes present in other *M. tuberculosis* strains but absent in H37Rv. Subject sequences with $\geq$98% identity to the query sequence at the amino acid level were selected and their known or predicted functions recorded. FASTQ sequence reads were also assembled against the annotated genomes of other strains of *M. tuberculosis* Euro-American Lineage 4, in particular CDC1551 (NC_002755), Erdman (NC_020559) and Haarlem (NC_022350). Lineage numbers refer to the Gagneux lineage classification system [16].

Confirmation of the presence of differentially encoded genes by PCR amplification and sanger re-sequencing

PCR was used to confirm the presence or absence of differentially encoded genes in the Rangipo and H37Rv strains. For this, pairs of oligonucleotides were designed using Primer 3 software (version 2.3.4) to amplify the full length and internal regions of the genes assayed (Supplementary Table S1). A consensus gene sequence, generated by aligning Rangipo genomes with the genome of MTBC strain CDC1551, was used as a template to design the primers. All pairs of primers were verified using the nucleotide BLASTN tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [15] to ensure that specific complementary regions were present only within the region of interest. For amplification reactions, a total reaction volume of 25 μL was used containing 0.5 μL DNA, 0.5 μL of 200 pM primer (forward and reverse each), 0.5 μL of 200 μM dNTPs (Invitrogen, Carlsbad, CA), 0.25 μL of 2U DNA polymerase (Invitrogen, Carlsbad, CA), 5 μL of 5 × buffer (Invitrogen, Carlsbad, CA) and 17.75 μL of ultrapure distilled water (Invitrogen, Carlsbad, CA) was used. The thermocycler settings were 98 °C for 30 seconds, followed by 30 cycles of 98 °C for 10 seconds, 55 °C for 30 seconds, 72 °C for 2 minutes, followed by 72 °C for 5 minutes and hold at 4 °C. The amplified fragments were separated by agarose gel electrophoresis and visualized using a Chemidoc XRS system (Bio-Rad Universal Hood II). The PCR products were excised from the agarose and column purified and Sanger sequencing of the products was performed at the Australian Genomic Research Facility (AGRF).

Phylogenetic tree building

The Rangipo genome sequences were assembled against the *M. tuberculosis* reference genome H37Rv to generate consensus sequences for each isolate. A maximum-likelihood phylogenetic tree was then built using the generalized time reversible (GTR) substitution model in PhyML [17] and included the annotated whole-genome sequences of the *M. tuberculosis* Lineage 4 strains

CDC1551 (NC_002755), H37Rv (NC_000962.3), Haarlem (NC_022350) and Erdman (NC_020559).

Results

Single-locus variations identified in the Rangipo strain of MTBC with respect to H37Rv

Single-locus variations (SLVs) were identified after aligning each Rangipo genome to H37Rv. A total of 727 SLVs were identified of which 700 polymorphisms were common to all 9 of the Rangipo isolates sequenced with respect to H37Rv. This compares to a previous report of 747 single nucleotide polymorphisms (SNP) with respect to H37Rv that were present in ten *M. tuberculosis* Rangipo isolates following genome assembly [18]. Of the common Rangipo SLVs, a total of 386 non-synonymous SLVs were identified which consisted of 354 missense variants, 21 frameshift mutations, 4 in-frame insertions, 4 stop codon gains, 2 in-frame deletions and 1 loss of stop codon mutation. The number of differences between individual Rangipo isolates ranged from 0 to 19 variant sites. A threshold of $\leq$5 SNPs between *M. tuberculosis* isolates has previously been proposed as an indicator of recent TB transmission between patients [19,20]. A number of the pairwise distances between isolates fall within the $\leq$5 SNP threshold (Table 3), but it should be noted that SNP distances between TB isolates can be influenced by factors such as time between patient sampling, local TB incidence and homogeneity of *M. tuberculosis* strains in some regions [21,22].

Non-synonymous variations in genes of known or predicted function

The genes containing the 386 non-synonymous SLVs were analysed with respect to their functional classification in terms of COGs [23]. Of these, 156 genes belonged to a single COG, with an additional 11 genes that could be assigned to two COGs. The remaining 219 genes were not assigned to an existing COG in the KEGG database (http://www.genome.jp/kegg/) [24] (Table 1). The most abundant COG among the genes containing the non-synonymous SLVs was COG C, energy production and conversion ($n = 16$) followed by COG L, replication, recombination and repair ($n = 15$). Conversely, the lowest numbers were classified under COG F, nucleotide transport and metabolism ($n = 1$); COG N, cell motility ($n = 2$); and COG D, cell cycle control, cell division and chromosome partitioning ($n = 3$). The Rangipo strains commonly contained variations with respect to H37Rv in genes determining metabolic function (59.5%) (Category: C, E, F, G, H, P, Q, R., S, I), followed by genes determining cellular process and signalling (23.5%) (Category: D, M, N, O, T, V), information storage and processing (16.8%) (Category: J, K, L) and phage- and transposon-associated proteins (1.63%) (Category: X). Rangipo common SLVs occurred most frequently in COG C (energy production and conversion; $n = 16$), COG L (replication, recombination and repair; $n = 15$) and COG H (coenzyme transport and metabolism; $n = 14$). It is important to note that COG categories C, H and L contain a relatively high number of genes, i.e. 102, 123, 81, respectively, which will likely impact on the frequency at which variants are detected in these categories.

Among the 700 SLVs common to the Rangipo strains sequenced, in-frame insertions were identified in four coding sequences (Rv0872c, Rv1435c, Rv2407 and Rv2823c) and deletions were detected in two coding sequences (Rv3345c and Rv0849). Gains of a stop codon were located in the genes *pstA*1 (Rv0930), *ipdA* (Rv3303), PE35 (Rv3872) and Rv0851c, and the loss of a stop codon was found in the polyketide beta-ketoacyl synthase *pks*3 gene (Rv1180) in all nine Rangipo isolates sequenced. In addition, there were 55 mutations occurring within PPE or PE_PGRS family proteins. Twelve non-

Table 1. Genes containing non-synonymous single-locus variations (nsSLVs), common to all nine Rangipo isolates sequenced, categorized with respect to clusters of orthologous groups (COGs).

COG	Functional category	Number of genes per COG
C	Energy production and conversion	16
D	Cell cycle control, cell division, chromosome partitioning	3
E	Amino acid transport and metabolism	12
F	Nucleotide transport and metabolism	1
G	Carbohydrate transport and metabolism	9
H	Coenzyme transport and metabolism	14
I	Lipid transport and metabolism	11
J	Translation, ribosomal structure and biogenesis	11
K	Transcription	4
L	Replication, recombination and repair	15
M	Cell wall/membrane/envelope biogenesis	13
N	Cell motility	2
O	Post translational modification, protein turnover, chaperones	9
P	Inorganic ion transport and metabolism	14
Q	Secondary metabolite biosynthesis, transport and catabolism	8
R	General function prediction only	9
S	Function unknown	12
T	Signal transduction mechanism	8
V	Defence mechanism	7
–	COG not designated	219
	Total	397

The genes containing the 386 non-synonymous SLVs, common to all nine Rangipo isolates sequenced, were analysed with respect to their functional classification in terms of clusters of orthologous group (COG). One hundred and fifty-six of the genes belonged to a single COG, with an additional 11 genes shared between two COGs. Two hundred and nineteen of the genes were not assigned to an existing COG in the KEGG database.

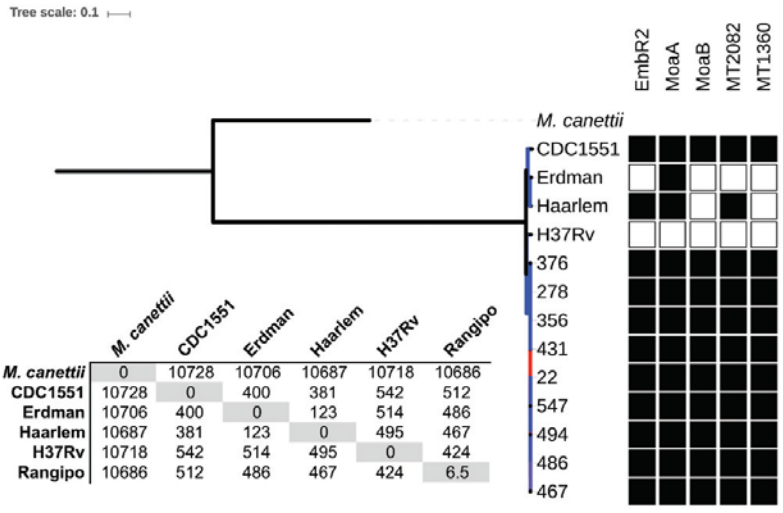
synonymous SLVs common to the Rangipo isolates were found in genes which have previously been associated with *M. tuberculosis* virulence, i.e. *aceAa*, *fadD28*, *kefB*, *mce1F*, *mycP1*, *pckA*, *pepD*, *phoR*, *pks15*, *plcB*, *pstA1*, *pstS1* [25].

Differentially encoded genes in the Rangipo strain

FASTQ sequences from the Rangipo strains, which did not map to the genome of H37Rv, were compared against the genome of other strains of *M. tuberculosis* by BLASTP analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [15]. This identified a number of genes ($n = 5$) which were present in the nine Rangipo isolates sequenced in this study and which had orthologues in strain CDC1551. The presence of the above five genes was confirmed in all nine of the Rangipo isolates by PCR amplification and subsequent Sanger sequencing and BLASTX analysis of the products (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [15]. The five genes were not amplified from the H37Rv template in agreement with their absence in the annotated genome sequence of this strain as reported by Fleischmann et al. [26]. The five Rangipo gene products consisted of orthologues of transcriptional regulator protein EmbR2 (MT3428), molybdopterin cofactor biosynthesis protein A (MT3427), molybdopterin cofactor biosynthesis protein B (MT3426), an extended-length DNA helicase (MT2082) and an additional adenylate cyclase (MT1360) from strain CDC1551. The closest orthologue of the helicase MT2082 in the H37Rv genome was Rv2024c (33% query cover, 99% identity at the nucleotide level). The closest orthologue of the adenylate cyclase MT1360 in the H37Rv genome was Rv1319c (99% query cover, 85% identity at the nucleotide level). The distribution of differentially encoded gene products in the *M. tuberculosis* Rangipo isolates compared to reference genomes of strains belonging to *M. tuberculosis* Lineage 4 is shown in Figure 1.

Phylogenetic analyses of the Rangipo strain

A maximum-likelihood-based distance tree was constructed in PhyML based on the consensus whole-genome sequences of the Rangipo isolates and the whole-genome sequences of the *M. tuberculosis* Lineage 4 strains CDC1551 (NC_002755), H37Rv (NC_000962.3), Haarlem (NC_022350) and Erdman (NC_020559), using the *M. canettii* to root the tree (Figure 1). This analysis



	M. canettii	CDC1551	Erdman	Haarlem	H37Rv	Rangipo
M. canettii	0	10728	10706	10687	10718	10686
CDC1551	10728	0	400	381	542	512
Erdman	10706	400	0	123	514	486
Haarlem	10687	381	123	0	495	467
H37Rv	10718	542	514	495	0	424
Rangipo	10686	512	486	467	424	6.5

Figure 1. Differential carriage of genes in the Rangipo strain and related *Mycobacterium tuberculosis* reference genomes. A maximum-likelihood phylogenetic tree was built using the generalized time reversible (GTR) substitution model in PhyML, including reference strains of *M. tuberculosis* Lineage 4. A total of 2,944,041 sites were identified on comparing Rangipo isolates with *M. tuberculosis* Lineage 4 strains CDC1551 (SRX393042), Haarlem (SRX347319) and Erdman (SRX364193) to the H37Rv (NC_000962.3) reference genome, while using the *Mycobacterium canettii* reference genome (HE572590.1) as a rooting. Branch lengths indicate divergence distance based on 11,438 common variant sites with bootstrap values indicated as follows: blue: 88–100, mauve: 73–87, red: <25 (of 100 replicates). The presence of genes *embR2*, *moaA*, *moaB*, *MT2082* and *MT1360* is indicated by filled squares to the right of the tree. A pairwise SNV distance matrix indicates strain divergence in terms of median observed variant sites.

Table 2. Non-synonymous single-locus variations, common to all nine Rangipo isolates sequenced, which occur in genes that have previously been related to mycobacterial virulence.

Locus tag	Gene	Role of gene in virulence	Nucleotide change	Amino acid change
Rv0174	*mce1F*	Growth regulation in murine macrophages	1109T > C	Leu370Pro
Rv0211	*pckA*	Growth regulation in human monocyte cell line	302A > C	Asn101Thr
Rv0758	*phoR*	Regulation of cell wall hydrophobicity	515C > T	Pro172Leu
Rv0930	*pstA1*	Resistance to host immunity	913C > T	Arg305[a]
Rv0934	*pstS1*	Adhesin binding to macrophages	63_64insA	Ala22[b]
Rv0983	*pepD*	Stress response protein	1169T > C	Leu390Pro
Rv1915	*aceAa*	Growth regulation in acetate containing medium	26A > C	Glu9Ala
Rv2350c	*plcB*	Growth regulation in macrophages	1406T > C	Leu469Ser
Rv2941	*fadD28*	Virulence attenuation in BALB/c mice	545T > C	Val182Ala
Rv2947c	*pks15*	Interaction with host cells	998T > C	Val333Ala
Rv3236c	*kefB*	Bacterial persistence in Guinea pig	962G > T	Arg321Leu
Rv3883c	*mycP1*	Regulation of ESX-1	979C > T	Pro327Ser

[a]Stop gained.
[b]Frameshift variant.

indicated that isolates 467, 486 and 494 are closely related in agreement with a low number of SLVs (0–1) between these isolates (Table. 3). Of the *M. tuberculosis* reference strains included in the analysis, the genome of H37Rv clustered more closely with the Rangipo isolates (Figure 1).

Table 3. Pairwise distance matrix of single-locus variations across the nine Rangipo isolates sequenced based on reference mapping to H37Rv.

	376	278	431	356	547	22	494	486	467
376		17	8	8	10	6	6	7	7
278			11	11	19	15	15	16	16
431				0	10	6	6	7	7
356					10	6	6	7	7
547						8	8	9	9
22							4	5	5
494								1	1
486									0
467									

Discussion

The association of the Rangipo genotype of *M. tuberculosis* with outbreaks of TB in New Zealand is well documented [4,5]. In this work, we investigated whether genomic data from the Rangipo genotype could inform us with respect to its inherent pathogenicity. Nine isolates of the genotype underwent whole-genome sequencing and comparative analysis was performed on the Rangipo sequences with respect to other virulent isolates of the organism.

Single-locus variant (SLV) analysis with respect to reference genome H37Rv revealed 700 polymorphisms that were common to all nine Rangipo isolates sequenced. Approximately 84.5% of these 700 polymorphisms occurred in coding regions and included 386 non-synonymous variants, 156 of which were located in genes which could be classified in one or more of the COGs (Table 1).

Twelve of the Rangipo common non-synonymous SLVs occurred in genes which have previously been reported to be important for *M. tuberculosis* pathogenesis, i.e. *aceAa*, *fadD28*, *kefB*, *mce1F*, *mycP1*, *pckA*, *pepD*, *phoR*, *pks15*, *plcB*, *pstA1*, *pstS1* (Table 2) [25]. Examples include the isocitrate lyase gene, *aceAa*, disruption of which has been associated with decreased maintenance of *M. tuberculosis* in the lungs and spleen of infected mice and reduced survival in murine macrophages and human blood monocyte-derived macrophages [27]. *M. tuberculosis* harbours four putative phospholipase C genes, *plcA-D*. Triple *plcABC* and quadruple *plcABCD* mutants are attenuated in terms of growth during the late phase of mouse infection [28]. The PhoPR two-component system is required for growth of *M. tuberculosis* under low Mg^{2+} availability *in vitro* and for growth during infection of macrophages and mice [29]. Similarly, mutation of the *pks15/1* polyketide synthase locus abrogates phenolic glycolipid production and results in decreased virulence in a rabbit model of meningeal tuberculosis [30].

While it is possible that one or more of the Rangipo common non-synonymous SLVs could affect the activity of the encoded products, experimental work with site-directed mutants would be needed to confirm the effect of specific variants on mycobacterial cell function and virulence. It is noteworthy that a single nucleotide polymorphism between *M. tuberculosis* strains CDC1551 and H37Rv causing a Leu152Pro substitution in the sensor kinase PhoR has been shown to increase cell wall hydrophobicity [31]. SLVs in the coding sequences of the *mce1* operon have been reported to be significantly high in clinical isolates of *M. tuberculosis* with *in silico* modelling predicting that a Pro359Ser substitution in Mce1A may have a diminishing effect on the stability of the protein and its biological function [32].

In addition to non-synonymous SLVs, our analysis also identified the differential carriage of genes that have been associated with the pathogenesis of *M. tuberculosis* in the Rangipo isolates with respect to reference strain H37Rv. Three virulence-related genes, which have orthologues in strain CDC1551, but which are absent in H37Rv, were identified in all nine Rangipo isolates whole-genome-sequenced in this study and were confirmed by PCR amplification and Sanger sequencing. These genes consisted of orthologues of transcriptional regulator protein EmbR2 (MT3428), molybdopterin cofactor biosynthesis protein A (MT3427) and molybdopterin cofactor biosynthesis protein B (MT3426) from strain CDC1551.

Molle et al. reported the presence in strain CDC1551 of EmbR2 (MT3428), an 1146 amino acid orthologue of the transcriptional regulator EmbR from H37Rv [33]. EmbR when phosphorylated by its cognate serine/threonine protein kinase, PknH, activates the transcription of the arabinosyltransferase genes, *embCAB*, which participate in the biosynthesis of arabinogalactans, key constituents of the mycobacterial cell wall. EmbR2 interacts with, but is not phosphorylated by, PknH [33]. Instead, EmbR2 inhibits the autokinase activity of PknH and the subsequent phosphoryl transfer to EmbR. A *pknH* mutant of *M. tuberculosis* has previously been shown to survive and replicate to a higher bacillary load in mouse lungs and spleens than its parental strain [34]. Therefore, EmbR2 through its control of the activity of the PknH/EmbR pair is believed to participate in the physiology and virulence of *M. tuberculosis* [33]. In our analysis, EmbR2 was present in the Rangipo isolates sequenced and in reference strains CDC1551 and Haarlem, but not in H37Rv or Erdman (Figure 1).

We also identified orthologues of the molybdopterin cofactor (MoCo) biosynthesis proteins MoaA (MT3427) and MoaB (MT3426) in the Rangipo isolates and the reference genome CDC1551 but not in H37Rv. MoaA participates in the conversion of guanosine triphosphate to cyclic pyranopterin monophosphate [35], while MoaB is involved in the adenylation of molybdopterin [36], key steps in MoCo biosynthesis. MoCo is an essential cofactor for redox reaction enzymes such as the *narGHI*-encoded nitrate reductase which functions in the adaptation of *M. tuberculosis* to hypoxic conditions [36], and is needed for the persistence of *M. tuberculosis* in the lungs of guinea pigs [37]. In addition, we identified an extended-length DNA helicase (MT2082, 1606 amino acids) and an additional adenylate cyclase (MT1360) in the Rangipo strain; however, their specific roles in virulence have yet to be elucidated.

Azhikina et al. previously reported the presence of orthologues of the five CDC1551 genes MT2082, MT3426, MT3427, MT3428 and MT1360 in clinical isolates of *M. tuberculosis* from Russia [38]. In addition, they also reported the presence of a truncated *plcD* gene which corresponded to the first 843 nucleotides of the 1545 base pair MT1799 gene of CDC1551. Interestingly, the Rangipo strain sequenced also exhibited a truncated *plcD* orthologue (905 nucleotides) with respect to MT1799.

Mycobacterium tuberculosis strain CDC1551 was first described in 1998 following a large TB outbreak in Tennessee and Kentucky between 1994 and 1996 [39]. The strain was notable for its high level of transmissibility from three patients to casual and close contacts in a community with a normally low risk for TB. Of 429 contacts, 311 (72.5%) produced positive tuberculin skin tests [39]. CDC1551 has been found to grow at a similar rate (doubling time of 25 h) as H37Rv (doubling time of 28 h) in the lungs of aerosol-infected mice during the first 14 days of infection [40]. But between day 14 and day 21, CDC1551 grew more slowly (doubling time of 105 h) than the H37Rv isolate used (doubling time of 36 h) [40]. Investigations in rabbits found that CDC1551 produced smaller granulomas containing fewer bacilli, which is indicative of lower virulence, as compared to H37Rv [41]. In addition, mice challenged with CDC1551 exhibited higher mRNA levels of inflammatory mediators TNF-α, IL-6, IL-10, IL-12 and IFN-γ, and a longer mean survival time of >250 days versus 185 days compared to H37Rv [40]. This has led to the suggestion that the original mini-epidemic caused by CDC1551 was due to higher transmissibility associated with the strain rather than increased in-host pathogenesis [41].

Conclusions

In summary, the Rangipo strain has been associated with high levels of transmissibility during community TB outbreaks. Our study identified key differences between the Rangipo strain and the standard *M. tuberculosis* reference strain, H37Rv that is commonly used for the assembly of whole-genome sequence reads. In addition to the presence of non-synonymous SLVs in genes linked to mycobacterial virulence, the Rangipo strain also harbours a number of virulence genes that are absent in H37Rv but present in TB outbreak strain, CDC1551. In vivo studies are needed to specifically correlate individual loci in the

Rangipo strain to its reported heightened transmissibility, and to understand the host response to infection by this strain.

Disclosure statement

The authors report no conflicts of interest.

Funding

RFOT is a recipient of a School of Medicine Research Development Grant, University of Tasmania and Royal Hobart Hospital Research Foundation grant (17-104). SSG is a recipient of a School of Medicine/Faculty of Health PhD Scholarship. Micheál Mac Aogáin was funded by Irish Research Council (EPSPD/2015/32).

ORCID

Micheál Mac Aogáin http://orcid.org/0000-0002-1726-7700
Ronan F. O'Toole http://orcid.org/0000-0002-4579-4479

References

[1] World Health Organization. Global tuberculosis report. 2015; Switzerland: World Health Organization.
[2] Institute of Environmental Science and Research Ltd (ESR). Tuberculosis in New Zealand: annual report 2014. 2015; New Zealand: Institute of Environmental Science and Research Ltd (ESR).
[3] Bissielo A, Lim E, Heffernan H. Tuberculosis in New Zealand: annual report 2010. New Zealand: Institute of Environmental Science and Research Ltd.
[4] De Zoysa R, Shoemack R, Vaughan R, et al. A prolonged outbreak of tuberculosis in the North Island. New Zealand Public Health Report. 2001.
[5] McElnay C, Thornley C, Armstrong R. A community and workplace outbreak of tuberculosis in Hawke's Bay in 2002. N Z Med J. 2004;117:U1019.
[6] Sexton K, Perera S, Pandey S. Five years of molecular typing of Mycobacterium tuberculosis isolates In New Zealand, 2003 to 2007. New Zealand: Institute of Environmental Science and Research Limited, 2007.
[7] Baussano I, Williams BG, Nunn P, et al. Tuberculosis incidence in prisons: a systematic review. PLoS Med. 2010;7:e1000381.
[8] Vinkeles Melchers NV, van Elsland SL, Lange JM, et al. State of affairs of tuberculosis in prison facilities: a systematic review of screening practices and recommendations for best TB control. PLoS One. 2013;8:e53644.
[9] Sacchi FP, Praca RM, Tatara MB, et al. Prisons as reservoir for community transmission of tuberculosis, Brazil. Emerging Infect Dis. 2015;21:452–455.
[10] Yen S, Bower JE, Freeman JT, et al. Phylogenetic lineages of tuberculosis isolates in New Zealand and their association with patient demographics. Int J Tuberc Lung Dis. 2013;17:892–897.
[11] Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537–544.
[12] Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–1649.
[13] Mac Aogain M, Gautam SS, Bower JE, et al. Draft Genome Sequence of a New Zealand Rangipo Strain of *Mycobacterium tuberculosis*. Genome Announc. 2016;4:e00657-16.
[14] Sekizuka T, Yamashita A, Murase Y, et al. TGS-TB: total genotyping solution for *Mycobacterium tuberculosis* using short-read whole-genome sequencing. PLoS One. 2015;10:e0142951.
[15] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol. 1990;215:403–410.
[16] Gagneux S, DeRiemer K, Van T, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA. 2006;103:2869–2873.
[17] Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
[18] Colangeli R, Arcus VL, Cursons RT, et al. Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. PLoS One 2014;9:e91024.
[19] Walker TM, Ip CLC, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis. 2013;13:137–146.
[20] Nikolayevskyy V, Kranzer K, Niemann S, et al. Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: a systematic review. Tuberculosis. 2016;98:77–85.
[21] Hatherell H-A, Colijn C, Stagg HR, et al. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC Med. 2016;14:21.
[22] Pouseele H, Supply P. Accurate whole-genome sequencing-based epidemiological surveillance of *Mycobacterium tuberculosis*. In: Sails A, Tang Y-W, editors. Methods in Microbiology. 42. USA: Academic Press; 2015.
[23] Galperin MY, Makarova KS, Wolf YI, et al. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43:D261–D2D9.
[24] Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353–D361.
[25] Forrellad MA, Klepp LI, Gioffré A, et al. Virulence factors of the *Mycobacterium tuberculosis* complex. Virulence. 2013;4:3–66.
[26] Fleischmann RD, Alland D, Eisen JA, et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol. 2002;184:5479–5490.
[27] Munoz-Elias EJ, McKinney JD. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. Nat Med. 2005;11:638–644.

[28] Raynaud C, Guilhot C, Rauzier J, et al. Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. Mol Microbiol. 2002;45:203–217.

[29] Walters SB, Dubnau E, Kolesnikova I, et al. The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol. 2006;60:312–330.

[30] Tsenova L, Ellison E, Harbacheuski R, et al. Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. J Infect Dis. 2005;192: 98–106.

[31] Schreuder LJ, Carroll P, Muwanguzi-Karugaba J, et al. *Mycobacterium tuberculosis* H37Rv has a single nucleotide polymorphism in PhoR which affects cell wall hydrophobicity and gene expression. Microbiology. 2015;161:765–773.

[32] Pasricha R, Chandolia A, Ponnan P, et al. Single nucleotide polymorphism in the genes of mce1 and mce4 operons of *Mycobacterium tuberculosis*: analysis of clinical isolates and standard reference strains. BMC Microbiol. 2011;11:41.

[33] Molle V, Reynolds Robert C, Alderwick LJ, et al. EmbR2, a structural homologue of EmbR, inhibits the *Mycobacterium tuberculosis* kinase/substrate pair PknH/EmbR. Biochem J. 2008;410:309–317.

[34] Papavinasasundaram KG, Chan B, Chung J-H, et al. Deletion of the *Mycobacterium tuberculosis* pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c Mice. J Bacteriol. 2005;187:5751–5760.

[35] Williams MJ, Kana BD, Mizrahi V. Functional analysis of molybdopterin biosynthesis in mycobacteria identifies a fused molybdopterin synthase in *Mycobacterium tuberculosis*. J Bacteriol. 2011;193:98–106.

[36] Williams M, Mizrahi V, Kana BD. Molybdenum cofactor: a key component of *Mycobacterium tuberculosis* pathogenesis? Crit Rev Microbiol. 2014;40:18–29.

[37] Williams MJ, Shanley CA, Zilavy A, et al. bis-Molybdopterin guanine dinucleotide is required for persistence of *Mycobacterium tuberculosis* in guinea pigs. Infect Immun. 2015;83:544–550.

[38] Azhikina T, Gvozdevsky N, Botvinnik A, et al. A genome-wide sequence-independent comparative analysis of insertion-deletion polymorphisms in multiple *Mycobacterium tuberculosis* strains. Res Microbiol. 2006;157:282–290.

[39] Valway SE, Sanchez MPC, Shinnick TF, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med. 1998;338: 633–639.

[40] Manca C, Tsenova L, Barry CE, 3rd, et al. *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. J Immunol. 1999;162:6740–6746.

[41] Bishai WR, Dannenberg AM, Jr., Parrish N, et al. Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. Infect Immun. 1999;67:4931–4934.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequence of a New Zealand Rangipo Strain of *Mycobacterium tuberculosis*

Micheál Mac Aogáin,[a] Sanjay S. Gautam,[b] James E. Bower,[c] Indira Basu,[c] Ronan F. O'Toole[a,b]

Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, St. James's Hospital, Dublin, Ireland[a]; Breathe Well Centre, School of Medicine, University of Tasmania, Hobart, Australia[b]; LabPlus, Auckland City Hospital, Auckland, New Zealand[c]

The Rangipo genotype of the *Mycobacterium tuberculosis* complex has been associated with a number of tuberculosis (TB) outbreaks in New Zealand. We report here the draft whole-genome sequence of a representative isolate of this strain.

Received 18 May 2016 **Accepted** 19 May 2016 **Published** 7 July 2016

Citation Mac Aogáin M, Gautam SS, Bower JE, Basu I, O'Toole RF. 2016. Draft genome sequence of a New Zealand Rangipo strain of *Mycobacterium tuberculosis*. Genome Announc 4(4):e00657-16. doi:10.1128/genomeA.00657-16.

Address correspondence to Ronan F. O'Toole, ronan.otoole@utas.edu.au.

Tuberculosis (TB) is a leading cause of infectious mortality worldwide, killing approximately 1.5 million people each year (1). In an earlier work, we determined that New Zealand isolates of the *Mycobacterium tuberculosis* complex were dominated by lineage 4 (Euro-American: 37.8%; 95% confidence interval [CI], 33.6 to 42.2%), followed by lineage 1 (Indo-Oceanic: 22.6%; 95% CI, 19.1 to 26.5%), lineage 2 (East Asian: 19.5%; 95% CI, 16.2 to 23.3%), and lineage 3 (East-African Indian: 17.7%; 95% CI, 14.5 to 21.3%) (2). In lineage 4 (Euro-American), 75 of the 184 isolates were in clusters (40.8%; 95% CI, 33.9 to 48.0%). The largest of these clusters belonged to the so-called Rangipo genotype based on 24-locus mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) typing.

In 2001, an analysis of an outbreak of TB in New Zealand, which occurred between November 1996 and May 2000, was reported (3). Forty-three of the 61 TB cases were confirmed to belong to the outbreak by IS*6110*-based restriction fragment length polymorphism (RFLP) typing of the isolates, with the remaining 18 cases determined by epidemiological contact tracing. One of the patients had served a prison sentence in the Tongariro/Rangipo Prison in 1998 and, as a result, the strain subsequently became referred to as the Rangipo strain (4).

Here, the genomic DNA of a Rangipo isolate, NZ494, was sequenced using an Illumina MiSeq instrument. A total of 2,128,439 paired-end reads were mapped to the *M. tuberculosis* strain H37Rv reference genome (accession no. AL123456/NC_000962) by the Burrows-Wheeler Aligner (5). This yielded an average read depth of 45-fold, covering 99.5% of the H37Rv genome. Variants relative to the H37Rv reference genome were called using the SAMtools analysis suite, and variant annotation was performed using SnpEff (6, 7). A 4,292,219-bp draft genome assembly of 179 contigs was assembled *de novo* using the SPAdes assembler (version 3.7) (8).

A total of 851 variant sites were identified relative to the H37Rv genome and consisted of 782 single-nucleotide variants (SNVs) and 69 insertions/deletions. Five hundred eighty-seven of the variants were nonsynonymous, of which 549 were SNVs and 38 were insertions/deletions. The genome of Rangipo isolate NZ494 did not display high-confidence single-nucleotide polymorphisms in genes correlating with antimicrobial drug resistance when analyzed using the PhyResSE database (9), which was consistent with the isolate's drug-susceptible phenotype.

An outbreak of TB in 2002 in Hawke's Bay involving 19 active cases of Rangipo TB was associated with a high rate of transmission, as determined by the presentation of TB disease or latent TB infection in 16.4% and 20.0%, respectively, of household or other close contacts (6). The reasons underlying the potentially high infectivity of the Rangipo strain are not known. Therefore, investigations on the virulence and antigenic determinants of the Rangipo genotype of *M. tuberculosis* are needed to generate a better understanding of its inherent propensity to transmit and cause disease in relation to other strains of the bacterium.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. LXWG00000000.

FUNDING INFORMATION

This work, including the efforts of Ronan F. O'Toole, was funded by University of Tasmania (UTAS) (School of Medicine Research Development Grant). This work, including the efforts of Sanjay S. Gautam, was funded by University of Tasmania (UTAS) (Faculty of Health/School of Medicine Ph.D. scholarship). This work, including the efforts of Micheál Mac Aogáin, was funded by Irish Research Council (EPSPD/2015/32).

REFERENCES

1. World Health Organization. 2015. Global tuberculosis report. World Health Organization, Geneva, Switzerland. http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf.
2. Yen S, Bower JE, Freeman JT, Basu I, O'Toole RF. 2013. Phylogenetic lineages of tuberculosis isolates in New Zealand and their association with patient demographics. Int J Tuberc Lung Dis 17:892–897. http://dx.doi.org/10.5588/ijtld.12.0795.
3. De Zoysa R, Shoemack R, Vaughan R, Vaughan A. 2001. A prolonged outbreak of tuberculosis in the North Island. N Z Public Health Rep 8:1–8.
4. McElnay C, Thornley C, Armstrong R. 2004. A community and workplace outbreak of tuberculosis in Hawke's Bay in 2002. N Z Med J 117:U1019.
5. Li H, Durbin R. 2010. Fast and accurate long-read alignment with

Burrows-Wheeler transform. Bioinformatics 26:589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
6. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92. http://dx.doi.org/10.4161/fly.19695.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
9. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, Cabibbe AM, Niemann S, Fellenberg K. 2015. PhyResSE: a Web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. J Clin Microbiol 53:1908–1914. http://dx.doi.org/10.1128/JCM.00025-15.

Genomics 109 (2017) 471–474

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Minireview

Limitations of the *Mycobacterium tuberculosis* reference genome H37Rv in the detection of virulence-related loci

Ronan F. O'Toole[a,b,*], Sanjay S. Gautam[a]

[a] School of Medicine, Faculty of Health, University of Tasmania, Hobart TAS 7000, Australia
[b] Department of Clinical Microbiology, Sir Patrick Dun Laboratory, School of Medicine, Trinity College Dublin, Dublin 8, Ireland

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Whole genome sequencing
H37Rv
CDC1551

ABSTRACT

The genome sequence of *Mycobacterium tuberculosis* strain H37Rv is an important and valuable reference point in the study of *M. tuberculosis* phylogeny, molecular epidemiology, and drug-resistance mutations. However, it is becoming apparent that use of H37Rv as a sole reference genome in analysing clinical isolates presents some limitations to fully investigating *M. tuberculosis* virulence. Here, we examine the presence of single locus variants and the absence of entire genes in H37Rv with respect to strains that are responsible for cases and outbreaks of tuberculosis. We discuss how these polymorphisms may affect phenotypic properties of H37Rv including pathogenicity. Based on our observations and those of other researchers, we propose that use of a single reference genome, H37Rv, is not sufficient for the detection and characterisation of *M. tuberculosis* virulence-related loci. We recommend incorporation of genome sequences of other reference strains, in particular, direct clinical isolates, in such analyses in addition to H37Rv.

1. Introduction

The first strain of *Mycobacterium tuberculosis* [1,2] to undergo whole-genome sequencing was H37Rv. The genome was sequenced by Cole and colleagues from the Pasteur Institute in Paris and the Wellcome Trust Sanger Centre in Cambridge, and published in 1998 in a landmark paper for the tuberculosis field [3]. H37Rv in itself is not a direct clinical isolate, but was derived in a laboratory from H37, the original strain collected from a patient at the Trudeau Institute, Saranac Lake, New York in 1905 [4]. Oatway and Steenken reported the dissociation of *M. tuberculosis* H37 into two variants, H37Rv and H37Ra, which exhibited high and low virulence, respectively, when inoculated through different routes into guinea pigs and rabbits [5]. H37Rv was found to exhibit a stippled, spreading colony with acid-fast rods whereas H37Ra was described as having a discrete crater-like worm-cast colony containing bacilli with granular and non-acid-fast tendencies [5]. A comparison of 39 strains of *M. tuberculosis* based on staining characteristics, cellular dimensions, colony morphology, growth rate, and animal pathogenicity, led to the deduction by Kubica and co-workers at the Trudeau Institute in 1972 that "the original tubercle bacilli of Koch are indeed those which today are referred to as *M. tuberculosis*, as exemplified by the neotype, H37Rv" [4].

2. H37Rv as a reference strain for TB research

The reported phenotypic similarity between H37Rv and the tuberculosis-causing bacteria described by Robert Koch in 1882 [6], combined with its worldwide distribution and common use, led to the adoption of H37Rv as a principal *M. tuberculosis* strain for tuberculosis vaccine [7], diagnostics [8], and drug therapy [9] research. With regard to genome-based studies, H37Rv has been used as the primary reference strain in multiple works on *M. tuberculosis* phylogeny, molecular epidemiology, and the detection of drug-resistance mutations [10–12]. Its genome is among the best curated of any bacterial species with a number of resources linking its gene annotations to transcriptomic, proteomic, and functional data such as Tuberculist (http://tuberculist.epfl.ch), KEGG (http://www.genome.jp/kegg/), and the TB Database (http://genome.tbdb.org/tbdb_sysbio/MultiHome.html). *M. tuberculosis* has been considered to be a highly monomorphic bacterium [13], hence choice of reference genome was not thought to have a significant effect on the genome analysis outputs. Indeed, based on a recent analysis of 162 closely-related Euro-American lineage 4 isolates using 7 different reference genomes, Lee and Behr reported that "the choice of reference genome, within the *M. tuberculosis* complex, has negligible influence on phylogeny and epidemiological studies of *M. tuberculosis* transmission"

Abbreviations: CFP-10, 10-kDa culture filtrate protein; ESAT-6, 6-kDa early secreted antigen target; ESX-1, ESAT-6 secretion system 1; HGC, homologous gene cluster; MoCo, molybdopterin cofactor; NADP, nicotinamide adenine dinucleotide phosphate; ORF, open-reading frame; PDIM, phthiocerol dimycocerosate; PE, proline-glutamate motif protein; PPE, proline-proline-glutamate motif protein

* Corresponding author at: School of Medicine, Faculty of Health, University of Tasmania, Medical Science 1 Building, 17 Liverpool Street, Hobart TAS 7000, Australia.
E-mail address: ronan.otoole@utas.edu.au (R.F. O'Toole).

http://dx.doi.org/10.1016/j.ygeno.2017.07.004
Received 25 June 2017; Received in revised form 17 July 2017; Accepted 18 July 2017
Available online 22 July 2017

[14]. H37Rv is a therefore a key tool in the study of *M. tuberculosis* and the preferred strain of choice for phylogenetic and epidemiological studies.

However, it is becoming more apparent from comparisons of the genomes of clinical isolates of *M. tuberculosis* that features of the H37Rv strain that can limit its robustness in the study of *M. tuberculosis* pathogenicity. For example, a previous study reported that H37Rv did not induce caseous necrosis in mice or the formation of multinucleate giant cells in contrast to clinical isolates of *M. tuberculosis* [15]. Other researchers have reported that low doses of H37Rv failed to produce clinically-recognisable disease in rhesus monkeys within 18 weeks of intra-bronchial inoculation, in contrast to the Erdman strain, although animals were tuberculin-skin test positive after 4 weeks [16]. In an *in vitro* infection assay using human monocyte-derived macrophages, H37Rv was found to exhibit a slower growth rate with respect to a patient isolate [17]. Due to the observed differences between strains, De Groote and co-workers have highlighted the need to use other strains of *M. tuberculosis*, in addition to H37Rv, when measuring the *in vivo* efficacy of anti-tubercular drug regimens [18]. Following a comparison of the virulence, immunopathology, and transmissibility of eight different clinical isolates of *M. tuberculosis* in mice, Marquina-Castillo and colleagues proposed that "that the current use of H37Rv as the standard for animal models may be flawed because there were important differences in pathology caused by H37Rv" [19].

3. Single-locus variations between H37Rv and other strains of *M. tuberculosis*

Prolonged laboratory culture of H37Rv, since isolation of H37 in 1905, could have potentially contributed to changes in this strain over time. Solans and co-investigators identified a + G insertion at the −74 position upstream of the *whiB6* (Rv3862c) start codon in the genomes of H37Rv and H37Ra that correlated with the decreased production of the 6-kDa early-secreted antigen target (ESAT-6) in these strains [20]. ESAT-6 is exported along with the 10-kDa culture filtrate protein (CFP-10) through the cell envelope of *M. tuberculosis* by the type VII ESAT-6 secretion system 1 (ESX-1) [21]. ESAT-6 mediates mycobacterial translocation from the phagosome to the cytosol which is required for the pathogenicity of *M. tuberculosis* [22]. The + G insertion upstream of the *whiB6* gene was not found in the genomes of 76 clinical isolates of *M. tuberculosis*, including CDC1551 which exhibits higher levels of ESAT-6 production with respect to H37Rv [20]. Introduction of a wild-type copy of the *whiB6* gene into H37Rv re-established ESAT-6 expression and secretion to the level of clinical strains highlighting the importance of the + G insertion [20].

The PhoP/PhoR two-component system is required for the growth of *M. tuberculosis* during infection of macrophages and mice [23]. The system regulates the biosynthesis of complex cell wall lipids *i.e.* sulphatides, diacyltrehalose, and polyacyltrehalose [23]. A single nucleotide polymorphism has been identified in H37Rv, with respect to CDC1551, which causes a leucine to proline substitution at position 152 of the sensor kinase PhoR (Rv0758) and is associated with elevated cell wall hydrophobicity [24]. Introduction of the Leu152Pro substitution into the CDC1551 background resulted in an increase in hydrophobicity to the level of the H37Rv strain indicating that the SNP has functional consequences for *M. tuberculosis* [24].

Comparisons have also been made between the genome sequences of H37Rv and its attenuated relative, H37Ra. These analyses have been useful for detecting possible H37Ra-specific variations that may contribute to its reduced virulence and include SNPs in genes that encode the two-component response regulator, PhoP, and the phospholipase C, PlcD, as well as members of the PE and PPE multi-protein families [25–27]. Applying the more recent single-molecule real time (SMRT) sequencing technology to re-sequencing of the H37Ra genome, Elghraoui et al. noted that the number of H37Ra-specific SNPs with respect to H37Rv was less than half that of the 76 previously-reported for the H37Ra genome sequence generated by Sanger-based shotgun sequencing of clones libraries [26,28]. The authors also reported that they identified an error at position 459399 of the H37Rv reference genome sequence upstream of Rv0383c. While this issue is not unique to H37Rv or H37Ra, the transition to newer sequencing platforms may involve revisiting some reference genomes obtained using earlier technologies.

4. Single-locus variations in different laboratory stocks of H37Rv

In addition to inter-strain single locus variations, Ioerger and colleagues performed whole-genome sequencing on isolates of H37Rv obtained from 6 laboratories in the USA and South Africa and detected genetic differences between different stocks of the strain [29]. Excluding differences shared with the genome sequence of strain H37Ra, the authors found 24 single nucleotide polymorphisms or indels among the six isolates of H37Rv, of which 22 occurred in coding regions. The variants are believed to have arisen from so-called *in vitro* evolution of the isolates following cultivation over time at the different sites [29]. Although the number of differences identified was not high (5–10 polymorphisms per strain), some of them may be functionally relevant to *M. tuberculosis*. For example, two of the isolates, H37RvLP and H37RvJO, were found to contain a + GC insertional frame-shift in the mycocerosic acid synthase gene, *mas* (Rv2940c), causing them to be deficient in biosynthesis of the cell wall glycolipid phthiocerol dimycocerosate (PDIM) [29]. An earlier signature-tagged mutagenesis study found that synthesis of PDIM is required for growth of *M. tuberculosis* in the lungs of mice [30]. Furthermore, Domenech and Reed reported that H37Rv is prone to losing its ability to produce PDIM following *in vitro* passage and recommended that researchers conducting *in vivo* virulence studies in this background strain confirm the presence of PDIM [31]. Owing to these differences among different stocks of H37Rv, Ioerger and co-workers concluded that "H37Rv as a standard reference strain should be used with some caution, as experimental results derived with 'H37Rv' may depend on the laboratory in which it is maintained and the associated unique genetic characteristics" [29].

5. Differences in gene content between H37Rv and *M. tuberculosis* clinical isolates

Fleischmann and colleagues were among the first to detect the absence of corresponding genes of strain *M. tuberculosis* CDC1551 in the H37Rv genome [32]. They noted the insertion of 17 complete open-reading frames (ORF) in CDC1551 that are not present in H37Rv. Among these ORFs, 9 have functional assignments that include an adenylate cyclase (MT1360), a glycosyl-transferase (MT1800), an oxidoreductase (MT1801), a 12 transmembrane-domain protein (MT1802), a membrane lipoprotein (MT2619), a proline-proline-glutamate motif (PPE) family protein (MT3248), paralogs of *moaB* (MT3426) and *moaA* (MT3427), and a putative transcription regulator (MT3428).

MoaA and MoaB are required for the biosynthesis of molybdopterin cofactor (MoCo), an essential co-factor for redox reaction enzymes such as the *narGHI*-encoded nitrate reductase [33,34]. Nitrate reductase has been shown to enable *M. tuberculosis*' adaptation to hypoxic conditions and its persistence in the lungs of guinea pigs [35]. Situated adjacent to the *moaA* and *moaB* genes, MT3428 encodes EmbR2, an orthologue of the transcriptional regulator EmbR (MT1305) [36]. EmbR2 inhibits the autokinase activity of PknH and the subsequent phosphoryl transfer to EmbR [36]. An *M. tuberculosis pknH* mutant has been shown to survive and replicate to a higher level in the lungs and spleens of mice than its parental strain [37]. Through its modulation of the kinase activity of PknH, EmbR2 is believed to participate in the physiology and virulence of *M. tuberculosis* [36].

From their analysis of 100 strains belonging to the *M. tuberculosis* complex, Brosch, Gordon and co-workers determined that the *moa* gene containing region, RvD5, was absent from strain H37Rv but present in

most of the *M. tuberculosis* clinical isolates analysed, including CDC1551, as well as in *M. africanum*, *M. bovis* and *M. microti* [38]. In addition to RvD5, H37Rv contains several other deletions mediated by homologous recombination between adjacent IS6110 elements [39,40]. Zheng and colleagues proposed a model for the evolution of strain H37Rv from a common ancestor of H37 and CDC1551 and the accumulation of deletions RvD1 to RvD6 in H37Rv that are not found in CDC1551 [26]. CDC1551 was first described in 1998 following a large TB outbreak in Tennessee and Kentucky between 1994 and 1996 and was noted for its high level of transmissibility to casual and close contacts in a community with a normally low risk for TB [41]. It is plausible that differences identified in gene content between H37Rv and CDC1551 play a role in the respective virulence of these strains.

In recent work on one of the major outbreak strains of *M. tuberculosis* in New Zealand, Rangipo, we used H37Rv as the reference genome. In addition to the traditional approach of assembling sequence reads against the H37Rv genome sequence, we examined unmapped reads using *de novo* assembly [42]. Additional ORFs were detected and the encoded protein products were predicted with blastp analysis [43]. This led to the identification of five genes, present exclusively in the Rangipo outbreak strain and not in H37Rv, that consisted of orthologues of the molybdopterin cofactor biosynthesis genes *moaA* (MT3427), *moaB* (MT3426), *embR2* transcription regulator (MT3428), adenylate cyclase (MT1360) and an extended-length DNA helicase (MT2082, 1606 amino acids) [42]. Applying a similar analysis to the genomes of a further 10 sequenced *M. tuberculosis* isolates from outbreaks described in Germany [44], Guatemala [45], Korea [46], and the UK [47] identified full-length homologous genes encoding MoaA and MoaB for all of the outbreak strains analysed. This further highlights the potential importance of these genes to the virulence of *M. tuberculosis*. In a genomic comparison involving a larger number of *M. tuberculosis* complex isolates ($n = 96$), Periwal and colleagues reported the complete absence of 10, and the partial absence of 64, homologous gene clusters (HGC) in H37Rv and H37Ra [48]. Among the HGCs that were validated as absent in H37Rv and H37Ra by polymerase-chain reaction analysis, genes with functional annotations include a transcriptional regulator (the oligonucleotides reported in the work map to the *embR2* gene, MT3428), a nicotinamide adenine dinucleotide phosphate (NADP)-dependent oxidoreductase, and an iron-regulated elongation factor (Tu), one or more of which may function in the virulence of *M. tuberculosis* [48].

The absence of a significant number of genes from H37Rv with respect to clinical isolates has led some researchers to develop a virtual *M. tuberculosis* consensus genome. Okumura and colleagues generated a virtual genome by merging the genome sequences of 19 *M. tuberculosis* complex strains [49]. The resultant virtual genome, 4,991,559 bp, was significantly larger than the genome of H37Rv, 4,411,532 bp (NC_000962.3), as it represented all of the sequence elements across the input genomes. It is possible that a virtual genome approach could facilitate pangenomic analysis of diverse strains of *M. tuberculosis* for core, accessory, and unique genes including genes that encode infectivity or immunity-related functions. But the main consideration is that reference genomes utilised in virulence studies need to encompass the fullest possible repertoire of genes present in newly-sequenced genomes of *M. tuberculosis*.

6. Conclusions

(i) H37Rv is an integral and highly-valued reference strain for studying the biology of *M. tuberculosis* including its phylogeny, molecular epidemiology, and drug-resistance.

(ii) The H37Rv genome exhibits a significant number of single locus polymorphisms as well as the absence of entire genes that impact the pathogenicity of this strain with respect to clinical isolates of *M. tuberculosis*.

(iii) The inclusion of further reference strains, in addition to H37Rv, in characterising the genome biology of *M. tuberculosis* is recommended, in particular, in relation to the virulence of the bacterium.

Conflict of interest

The authors declare no financial or other conflicting interests.

Acknowledgements

The authors would like to acknowledge the support of the Royal Hobart Hospital Research Foundation (grant #17-104).

Authors' contributions

RFOT conceived the minireview topic and contributed to drafting and reviewing the manuscript. SSG contributed to the drafting of the manuscript.

References

[1] K.B. Lehmann, R.O. Neumann, Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik, 1 (1896), pp. 1–448.
[2] V.B.D. Skerman, V. McGowan, P.H.A. Sneath, Approved Lists of Bacterial Names (Amended), ASM Press, Washington (DC), 1989.
[3] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, et al., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[4] G.P. Kubica, T.H. Kim, F.P. Dunbar, Designation of strain H37Rv as the neotype of *Mycobacterium tuberculosis*, Int. J. Syst. Evol. Microbiol. 22 (1972) 99–106.
[5] W.H. Oatway, W. Steenken, The pathogenesis and fate of tubercle produced by dissociated variants of tubercle bacilli, J Infect Dis 59 (1936) 306–325.
[6] R. Koch, Die Aetiologie der Tuberkulose, Berl Klin Wchschr 19 (1882) 221.
[7] G. Kaplan, Rational vaccine development—a new trend in tuberculosis control, N. Engl. J. Med. 353 (2005) 1624–1625.
[8] S.L. Zhang, J.W. Zhao, Z.Q. Sun, E.Z. Yang, J.H. Yan, Q. Zhao, et al., Development and evaluation of a novel multiple-antigen ELISA for serodiagnosis of tuberculosis, Tuberculosis (Edinb.) 89 (2009) 278–284.
[9] N.K. Dutta, A. Alsultan, T.J. Gniadek, D.A. Belchis, M.L. Pinn, K.E. Mdluli, et al., Potent rifamycin-sparing regimen cures guinea pig tuberculosis as rapidly as the standard regimen, Antimicrob. Agents Chemother. 57 (2013) 3910–3916.
[10] T.M. Walker, T.A. Kohl, S.V. Omar, J. Hedge, Elias C. Del Ojo, P. Bradley, et al., Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study, Lancet Infect. Dis. 15 (2015) 1193–1202.
[11] A. Roetzer, R. Diel, T.A. Kohl, C. Ruckert, U. Nubel, J. Blom, et al., Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study, PLoS Med. 10 (2013) e1001387.
[12] O. Mestre, T. Luo, T. Dos Vultos, K. Kremer, A. Murray, A. Namouchi, et al., Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair, PLoS ONE 6 (2011) e16020.
[13] M. Achtman, Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens, Annu. Rev. Microbiol. 62 (2008) 53–70.
[14] R.S. Lee, M.A. Behr, Does choice matter? Reference-based alignment for molecular epidemiology of tuberculosis, J. Clin. Microbiol. 54 (2016) 1891–1895.
[15] B. Al Shammari, T. Shiomi, L. Tezera, M.K. Bielecka, V. Workman, T. Sathyamoorthy, et al., The extracellular matrix regulates granuloma necrosis in tuberculosis, J Infect Dis 212 (2015) 463–473.
[16] B.J. Gormus, J.L. Blanchard, X.H. Alvarez, P.J. Didier, Evidence for a rhesus monkey model of asymptomatic tuberculosis, J. Med. Primatol. 33 (2004) 134–145.
[17] M. Zhang, J. Gong, Y. Lin, P.F. Barnes, Growth of virulent and avirulent *Mycobacterium tuberculosis* strains in human macrophages, Infect. Immun. 66 (1998) 794–799.
[18] M.A. De Groote, V. Gruppo, L.K. Woolhiser, I.M. Orme, J.C. Gilliland, A.J. Lenaerts, Importance of confirming data on the in vivo efficacy of novel antibacterial drug regimens against various strains of *Mycobacterium tuberculosis*, Antimicrob. Agents Chemother. 56 (2012) 731–738.
[19] B. Marquina-Castillo, L. Garcia-Garcia, A. Ponce-de-Leon, M.E. Jimenez-Corona, M. Bobadilla-Del Valle, B. Cano-Arellano, et al., Virulence, immunopathology and transmissibility of selected strains of *Mycobacterium tuberculosis* in a murine model, Immunology 128 (2009) 123–133.
[20] L. Solans, N. Aguiló, S. Samper, A. Pawlik, W. Frigui, C. Martín, et al., A specific polymorphism in *Mycobacterium tuberculosis* H37Rv causes differential ESAT-6 expression and identifies WhiB6 as a novel ESX-1 component, Infect. Immun. 82 (2014) 3446–3456.
[21] D. Houben, C. Demangel, J. van Ingen, J. Perez, L. Baldeón, A.M. Abdallah, et al., ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria,

Cell. Microbiol. 14 (2012) 1287–1298.
[22] S.V. Jamwal, P. Mehrotra, A. Singh, Z. Siddiqui, A. Basu, K.V.S. Rao, Mycobacterial escape from macrophage phagosomes to the cytoplasm represents an alternate adaptation mechanism, Sci Rep 6 (2016) 23089.
[23] S.B. Walters, E. Dubnau, I. Kolesnikova, F. Laval, M. Daffe, I. Smith, The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis, Mol. Microbiol. 60 (2006) 312–330.
[24] L.J. Schreuder, P. Carroll, J. Muwanguzi-Karugaba, R. Kokoczka, A.C. Brown, T. Parish, *Mycobacterium tuberculosis* H37Rv has a single nucleotide polymorphism in PhoR which affects cell wall hydrophobicity and gene expression, Microbiology 161 (2015) 765–773.
[25] S. Kohli, Y. Singh, K. Sharma, A. Mittal, N.Z. Ehtesham, S.E. Hasnain, Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H(3)(7)Rv and H(3)(7)Ra reveal novel and interesting differences with implications in virulence, Nucleic Acids Res. 40 (2012) 7113–7122.
[26] H. Zheng, L. Lu, B. Wang, S. Pu, X. Zhang, G. Zhu, et al., Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv, PLoS ONE 3 (2008) e2375.
[27] X. Jia, L. Yang, M. Dong, S. Chen, L. Lv, D. Cao, et al., The bioinformatics analysis of comparative genomics of *Mycobacterium tuberculosis* complex (MTBC) provides insight into dissimilarities between intraspecific groups differing in host association, virulence, and epitope diversity, Front. Cell. Infect. Microbiol. 7 (2017) 88.
[28] A. Elghraoui, S.J. Modlin, F. Valafar, SMRT genome assembly corrects reference errors, resolving the genetic basis of virulence in *Mycobacterium tuberculosis*, BMC Genomics 18 (2017) 302.
[29] T.R. Ioerger, Y. Feng, K. Ganesula, X. Chen, K.M. Dobos, S. Fortune, et al., Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories, J. Bacteriol. 192 (2010) 3645–3653.
[30] J.S. Cox, B. Chen, M. McNeil, W.R. Jacobs Jr., Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice, Nature 402 (1999) 79–83.
[31] P. Domenech, M.B. Reed, Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown in vitro: implications for virulence studies, Microbiology 155 (2009) 3532–3543.
[32] R.D. Fleischmann, D. Alland, J.A. Eisen, L. Carpenter, O. White, J. Peterson, et al., Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains, J. Bacteriol. 184 (2002) 5479–5490.
[33] M.J. Williams, B.D. Kana, V. Mizrahi, Functional analysis of molybdopterin biosynthesis in mycobacteria identifies a fused molybdopterin synthase in *Mycobacterium tuberculosis*, J. Bacteriol. 193 (2011) 98–106.
[34] M. Williams, V. Mizrahi, B.D. Kana, Molybdenum cofactor: a key component of *Mycobacterium tuberculosis* pathogenesis? Crit. Rev. Microbiol. 40 (2014) 18–29.
[35] M.J. Williams, C.A. Shanley, A. Zilavy, B. Peixoto, C. Manca, G. Kaplan, et al., bis-Molybdopterin guanine dinucleotide is required for persistence of *Mycobacterium tuberculosis* in guinea pigs, Infect. Immun. 83 (2015) 544–550.
[36] V. Molle, C. Reynolds Robert, J. Alderwick Luke, S. Besra Gurdyal, J. Cozzone Alain, K. Fütterer, et al., EmbR2, a structural homologue of EmbR, inhibits the *Mycobacterium tuberculosis* kinase/substrate pair PknH/EmbR, Biochem. J. 410 (2008) 309–317.
[37] K.G. Papavinasasundaram, B. Chan, J.-H. Chung, M.J. Colston, E.O. Davis, Y. Av-Gay, Deletion of the *Mycobacterium tuberculosis* pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice, J. Bacteriol. 187 (2005) 5751–5760.
[38] R. Brosch, S.V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, et al., A new evolutionary scenario for the *Mycobacterium tuberculosis* complex, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 3684–3689.
[39] S.V. Gordon, R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, S.T. Cole, Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays, Mol. Microbiol. 32 (1999) 643–655.
[40] R. Brosch, W.J. Philipp, E. Stavropoulos, M.J. Colston, S.T. Cole, S.V. Gordon, Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain, Infect. Immun. 67 (1999) 5768–5774.
[41] S.E. Valway, M.P.C. Sanchez, T.F. Shinnick, I. Orme, T. Agerton, D. Hoy, et al., An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*, N. Engl. J. Med. 338 (1998) 633–639.
[42] S.S. Gautam, M. Mac Aogain, J.E. Bower, I. Basu, R.F. O'Toole, Differential carriage of virulence-associated loci in the New Zealand Rangipo outbreak strain of *Mycobacterium tuberculosis*, Infect. Dis. 49 (2017) 680–688 (Lond).
[43] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[44] A. Roetzer, S. Schuback, R. Diel, F. Gasau, T. Ubben, A. di Nauta, et al., Evaluation of *Mycobacterium tuberculosis* typing methods in a 4-year study in Schleswig-Holstein, Northern Germany, J. Clin. Microbiol. 49 (2011) 4173–4178.
[45] J.W. Saelens, D. Lau-Bonilla, A. Moller, N. Medina, B. Guzmán, M. Calderón, et al., Whole genome sequencing identifies circulating Beijing-lineage *Mycobacterium tuberculosis* strains in Guatemala and an associated urban outbreak, Tuberculosis (Edinb.) 95 (2015) 810–816.
[46] S.J. Han, T. Song, Y.-J. Cho, J.-S. Kim, S.Y. Choi, H.-E. Bang, et al., Complete genome sequence of *Mycobacterium tuberculosis* K from a Korean high school outbreak, belonging to the Beijing family, Stand. Genomic Sci. 10 (2015) 78.
[47] M.E. Török, S. Reuter, J. Bryant, C.U. Köser, S.V. Stinchcombe, B. Nazareth, et al., Rapid whole-genome sequencing for investigation of a suspected tuberculosis outbreak, J. Clin. Microbiol. 51 (2013) 611–614.
[48] V. Periwal, A. Patowary, S.K. Vellarikkal, A. Gupta, M. Singh, A. Mittal, et al., Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome, PLoS ONE 10 (2015) e0122979.
[49] K. Okumura, M. Kato, T. Kirikae, M. Kayano, T. Miyoshi-Akiyama, Construction of a virtual *Mycobacterium tuberculosis* consensus genome and its application to data from a next generation sequencer, BMC Genomics 16 (2015) 218.

PLOS ONE

RESEARCH ARTICLE

Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case

Sanjay S. Gautam[1], Micheál Mac Aogáin[2], Louise A. Cooley[3], Greg Haug[4], Janet A. Fyfe[5], Maria Globan[5], Ronan F. O'Toole[1,2]*

1 School of Medicine, University of Tasmania, Hobart, Tasmania, Australia, **2** Department of Clinical Microbiology, Trinity College Dublin, Dublin, Ireland, **3** Royal Hobart Hospital, Hobart, Tasmania, Australia, **4** Launceston General Hospital, Launceston, Tasmania, Australia, **5** Mycobacterium Reference Laboratory, Victoria Infectious Diseases Reference Laboratory, Peter Doherty Institute, Melbourne, Victoria, Australia

* ronan.otoole@utas.edu.au

OPEN ACCESS

Citation: Gautam SS, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, Globan M, et al. (2018) Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. PLoS ONE 13(2): e0192351. https://doi.org/10.1371/journal.pone.0192351

Editor: Seyed Ehtesham Hasnain, Indian Institute of Technology Delhi, INDIA

Received: October 26, 2017

Accepted: January 22, 2018

Published: February 21, 2018

Data Availability Statement: All relevant data are within the paper.

Funding: This research was supported by funding from the Royal Hobart Hospital Research Foundation, grant reference number 17-104.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Background

The origin and spread of tuberculosis (TB) in Tasmania and the types of strains of *Mycobacterium tuberculosis* complex (MTBC) present in the population are largely unknown.

Objective

The aim of this study was to perform the first genomic analysis of MTBC isolates from Tasmania to better understand the epidemiology of TB in the state.

Methods

Whole-genome sequencing was performed on cultured isolates of MTBC collected from 2014–2016. Single-locus variant analysis was applied to determine the phylogeny of the isolates and the presence of drug-resistance mutations. The genomic data were then cross-referenced against public health surveillance records on each of the cases.

Results

We determined that 83.3% of TB cases in Tasmania from 2014–2016 occurred in non-Australian born individuals. Two possible TB clusters were identified based on single locus variant analysis, one from November-December 2014 ($n = 2$), with the second from May-August 2015 ($n = 4$). We report here the first known isolate of multi-drug resistant (MDR) *M. tuberculosis* in Tasmania from 2016 for which we established its drug resistance mutations and potential overseas origin. In addition, we characterised a case of *M. bovis* TB in a Tasmanian-born person who presented in 2014, approximately 40 years after the last confirmed case in the state's bovids.

Conclusions

TB in Tasmania is predominantly of overseas origin with genotypically-unique drug-susceptible isolates of *M. tuberculosis*. However, the state also exhibits features of TB that are observed in other jurisdictions, namely, the clustering of cases, and drug resistance. Early detection of TB and contact tracing, particularly of overseas-born cases, coordinated with rapid laboratory drug-susceptibility testing and molecular typing, will be essential for Tasmania to reach the World Health Organisation's TB eradication goals for low-incidence settings.

Introduction

There are few descriptions in the literature describing the epidemiology of tuberculosis (TB) in Tasmania or the types of strains of *Mycobacterium tuberculosis* present in its population. Tasmania is a small island state in Australia with approximately 0.5 million people [1]. European settlement of its capital, Hobart, began when it was founded as a penal colony in 1803 by Colonel David Collins who came to Australia with the First Fleet in 1788. The earliest documented evidence of TB in Tasmania comes from Collins who reported that among 36 ill people in his Hobart settlement in 1804, one had consumption [2]. The first recorded deaths from TB in the Tasmanian Aboriginal population occurred between 1835 and 1838 on Flinders Island, situated to the north east of the main Tasmanian island, where they had been relocated from 1830 [2]. The burden of TB in the state increased until reaching a peak of 248 cases in 1940 that corresponded to an incidence rate of 103.8/100,000 in Tasmania compared to the national incidence rate of 59.3/100,000 in Australia at that time [3].

Today, Tasmania is considered a low TB-burden state with an incidence rate of 1.7/100,000 persons compared to 5.7/100,000 nationally in 2014 [4]. There are a number of features of TB in the state that are of interest. Firstly, TB in Tasmania has been considered to consist of isolated unique cases that have been imported from other jurisdictions. Secondly, the state has been free of multi-drug resistant forms of TB. Thirdly, as part of the Brucellosis and Tuberculosis Eradication Campaign (BTEC) in Australia, bovine TB disease was eradicated from Tasmanian cattle herds in 1975 thus, eliminating the primary source of human cases of *M. bovis* TB in the state [5].

In this study, we performed an in-depth analysis of the types of MTBC strains isolated in Tasmania between 2014 and 2016 using whole-genome sequencing. We then correlated genomic information with public health surveillance data to better define the epidemiology of TB in Tasmania.

Methods

Study design

Samples from 18 cultured isolates collected in Tasmania from 2014 to 2016, inclusive, were available for this study. This represents 62.1% of total TB notifications that occurred during this time period (n = 29) [6]. Diagnostic laboratory and clinical data were obtained from the Royal Hobart Hospital, Launceston General Hospital, and the Victorian Infectious Diseases Reference Laboratory (VIDRL). Samples were sent to the School of Medicine, University of Tasmania, for next generation sequencing and whole-genome bioinformatics analysis. Ethics approval for this study was obtained from the Tasmanian Health and Medical Human

Research Ethics Committee (H0016214). A waiver of consent was acquired as the study was an observational non-interventional analysis of MTBC isolates and de-identified data that were obtained from routine laboratory testing.

Sample processing, culture, and drug susceptibility testing

Specimens from patients suspected of having tuberculosis were cultured using both solid (Brown and Buckle agar, Löwenstein-Jensen agar) and liquid media (Mycobacterial Growth Indicator Tubes (MGIT)) in accordance with standard protocols for mycobacterial growth [7]. Ziehl-Neelsen staining, TB MPT64 antigen test (Standard Diagnostics Bioline TB MPT64 antigen test) and TB PCR (GenXpert, Cepheid) were performed on positive cultures. The isolates were supplied to a reference laboratory for further characterisation. Drug-susceptibility testing was performed using the MGIT system [7].

Genomic DNA isolation

1.5 mL of heat-inactivated mycobacterial cultures were centrifuged at 8,000 rpm for 3 minutes at room temperature. The cell pellet was resuspended in 200 μL phosphate-buffered saline and treated with 25 μL of 10 mg/mL lysozyme and incubated at 37˚C for 1 hour followed by 95˚C for 15 minutes. 30 μL proteinase K (10 mg/mL) were added and the sample was incubated at 55˚C for 30 minutes. A Qiagen DNeasy Blood and Tissue kit was then used to extract mycobacterial genomic DNA as per the manufacturer's instructions and the DNA was eluted with 200 μL of Buffer AE. 1 μL of RNase A (7000 units/mL, Qiagen) was added to 50 μL of genomic DNA eluent and incubated at room temperature for 1 hour. The genomic DNA was further purified using a High Pure PCR Template Preparation Kit as per the manufacturer's instructions (Roche) and quantified using the Quant-iT Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific).

Whole genome sequencing and data analysis

Purified genomic DNA was tagged and amplified using a Nextera® XT DNA Library Preparation Kit and Nextera® XT Index Kit as per the manufacturer's (Illumina) instructions. The libraries generated were cleaned using Agencourt AMPure XP beads, normalized and then pooled. The concentration of the pooled library was determined by qPCR using a KAPA Library Quantification Kit. 15 pM of the pooled library were loaded into a MiSeq Reagent Kit v2 cartridge and run on an Illumina MiSeq instrument generating paired-end reads of 150 base pairs (bp) (maximum). The fastq sequence files were collected and analysed using Geneious software suite (R 9.5) [8]. Paired-end reads were trimmed (error probability limit of 0.05) and then mapped (random multiple base matches) to the publicly-available annotated genome of *M. tuberculosis* reference strain H37Rv (accession number NC_000962.3) [9] using a maximum variant p-value of 10^{-6} when exceeding 65% bias. Single-locus variations (SLV) were called at a minimum variant frequency of 95% and a minimum mean genome coverage of 20, and were annotated as previously described [10]. Mycobacterial lineage was predicted with TB Profiler [11] and mutations associated with drug resistance were detected using the PhyResSE database [12] followed by manual checking of the sequence. A maximum likelihood phylogenetic inference tree was built in PhyML using the generalised time reversible (GTR) substitution model [13].

TB epidemiological analysis

A threshold of ≤5 single nucleotide polymorphisms (SNPs) between *M. tuberculosis* isolates has previously been proposed as an indicator of recent TB transmission between patients, while >12 SNP differences between isolates has been considered as evidence against recent transmission [14–16]. SNP distances between TB isolates can be affected by factors such as time between patient sampling, local TB incidence, and homogeneity of *M. tuberculosis* strains in some regions [17, 18]. Therefore, for isolates that were within 5 SNP differences of one another, additional epidemiological data were used. The definition of a possible cluster was based on the National Tuberculosis Advisory Committee of Australia's guidelines which state that "A 'possible cluster' will be any 2 or more active cases with the same genotype as defined by the method used where temporal and geospatial association is plausible but no direct epidemiological link is identified" [19]. Each *M. tuberculosis* isolate in this work was characterised based on SNP differences to other members of the same global lineage, and the presence of spatiotemporal links between cases. Furthermore, *in silico* spoligotyping of the isolates was performed using the Total Genotyping Solution for TB (TGS-TB) database [20] and compared with published data on MTBC genotypes in the patient's country/region of origin.

Results

Relative distribution of MTBC lineages in Tasmania

Cases were 72.2% male (n = 13), 27.8% female (n = 5) (Table 1). The age of patients at date of specimen collection ranged from 3 months to 70 years of age with a mean age of 33.6 years. 77.7% (n = 14) of cases were pulmonary and 22.2 (n = 4) were extra-pulmonary. 83.3% of cases (n = 15) were non-Australian born and 16.7% (n = 3) were Australian-born (Table 1).

Table 1. Demographic and specimen information for tuberculosis cases (n = 18) in Tasmania from 2014 to 2016. Demographic variables on the TB cases and specimen types were recorded. Cases were 72.2% male and 27.8% female. The mean TB patient age was 33.6 years (range 0–70 years).

Isolate Name	Age Range of Patient (years)	Specimen Type	Year of Specimen Collection	MTBC Lineage	Patient Country of Origin
RHH2	20–39	Sputum	2015	1	Philippines
RHH3	≥60	Sputum	2015	3	Nepal
RIII14	20–39	Sputum	2016	4	Thailand
RIII15	20–39	Paraspinal aspirate	2015	1	Myanmar/Malaysia
RIII16	20–39	Osteomyelitis	2014	3	Nepal
RHH7	40–59	Sputum	2014	4	New Zealand
RHH8	20–39	Sputum	2016	2	Malaysia
RHH9	<5	Gastric aspirate	2014	4	New Zealand
RHH10	20–39	Sputum	2016	3	Nepal
RIII11	40–59	Sputum	2015	3	Nepal
RIII12	<5	Gastric aspirate	2016	1	Philippines
RHH13	20–39	Sputum	2015	3	Nepal
RIII14	<5	Gastric aspirate	2015	3	Nepal
RIII15	>60	Sputum	2014	4	Australia
TTB1	>60	Urine	2016	*M. bovis* BCG	Australia
TASMDR1	20–39	Tissue	2016	2	Vietnam
TTB3	20–39	Sputum	2016	1	Philippines
TASMB14	≥60	Sputum	2014	*M. bovis*	Australia

MTBC, *Mycobacterium tuberculosis* complex

https://doi.org/10.1371/journal.pone.0192351.t001

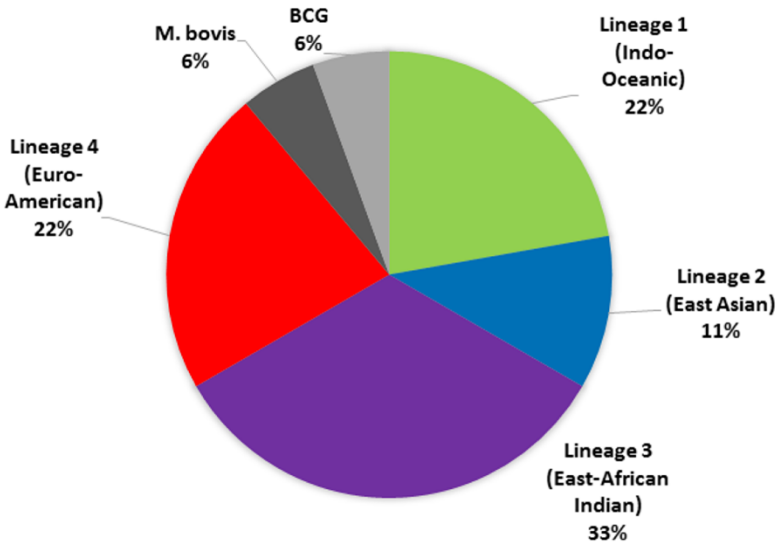


Fig 1. Relative frequency of *Mycobacterium tuberculosis* complex (n = 18) isolates in Tasmania from 2014 to 2016.

https://doi.org/10.1371/journal.pone.0192351.g001

Whole-genome sequence data were obtained for the 18 TB isolates analysed from 2014–2016. The phylogenetic lineage of each isolate was determined using the PhyResSE and TB Profiler databases [11, 12]. The most common lineage among the samples analysed was the East-African Indian Lineage 3 (n = 6, 33.3%) followed by the Euro-American Lineage 4 (n = 4, 22.2%), Indo-Oceanic Lineage 1 (n = 4, 22.2%), and East-Asian Lineage 2 (n = 2, 11.1%) (Fig 1). In addition, cases of TB due to *M. bovis* (n = 1) and *M. bovis* BCG (n = 1) were recorded in 2014 and 2016, respectively.

Molecular epidemiological clustering of TB cases

A maximum likelihood phylogenetic inference tree generated using PhyML revealed grouping of the different isolates into specific clades which were in agreement with the lineage analysis performed using the PhyResSE and TB Profiler databases (Fig 2). In addition, the phylogenetic tree revealed possible genetic clusters of isolates within Lineages 3 and 4.

The Lineage 3 cluster consisted of four isolates (RHH3, RHH11, RHH13, and RHH14) which exhibited zero SLV differences with respect to one another. Based on the proposed threshold of ≤5 SNPs, from Walker and others [14, 15], this is indicative of recent transmission. The isolates were collected in Tasmania from May to August 2015 from drug-susceptible cases of pulmonary TB. The patients were household contacts and were originally from Nepal. A previous analysis of 261 *M. tuberculosis* isolates collected in Nepal from pulmonary TB patients between August 2009 and August 2010 using spoligotyping and real-time PCR analysis of SNPs found that the most frequent *M. tuberculosis* lineage was Lineage 3 (40.6%) [21]. The *in silico* spoligotype of all four Tasmanian Lineage 3 cluster isolates matched Spoligotype International Type 26 of the CAS1_Delhi spoligotyping family which was found to account for approximately 50% of Lineage 3 *M. tuberculosis* isolates in the previous analysis of Nepal TB cases (Fig 3) [21].

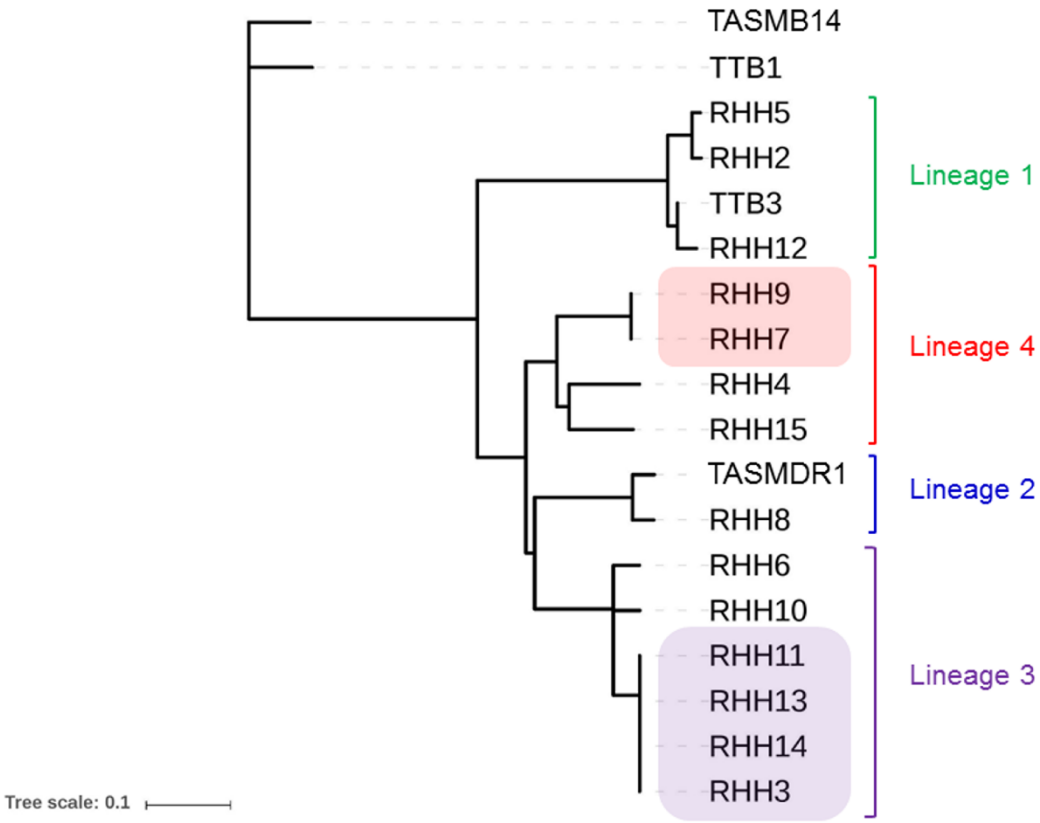


Fig 2. Phylogenetic relationship of *Mycobacterium tuberculosis* complex isolates in Tasmania from 2014 to 2016. TASMB14 and TTB1 constitute isolates of *M. bovis* and *M. bovis* BCG, respectively. The Gagneux lineage numbers are indicated for the other isolates. Lineage 2 isolate TASMDR1 is a multi-drug resistant isolate of *Mycobacterium tuberculosis*. Lineage 3 isolates RHH3, 11, 13 and 14 are identical (zero SNP differences with respect to one another) and form an epidemiological cluster as do Lineage 4 isolates RHH7 and RHH9. The phylogenetic tree was built using PhyML (Generalised Time Reversible substitution model).

https://doi.org/10.1371/journal.pone.0192351.g002

The Lineage 4 cluster comprised two isolates (RHH7 and RHH9) which exhibited zero SLV differences with respect to one another. The isolates were collected in Tasmania from November to December 2014 from drug-susceptible cases of pulmonary TB. The patients were household contacts and were originally from New Zealand. From an earlier study involving 487 MTBC isolates collected in New Zealand from January 2010 to December 2011, the most prevalent lineage was Lineage 4 which made up 37.8% of TB cases in the general population and 70.5% of cases in the New Zealand-born population [23].

The remaining MTBC isolates (n = 12) were phylogenetically unique with respect to one another with >50 SNP differences between them. The closest isolates from this group were isolates TTB3 and RHH12 which differ by 74 single-locus variations and belong to Indo-Oceanic Lineage 1. We do not have epidemiological evidence that these patients formed a probable cluster. Both patients were originally from the Philippines where Lineage 1 is highly dominant among TB isolates [24, 25].

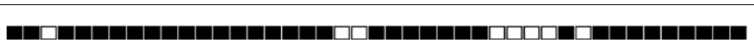
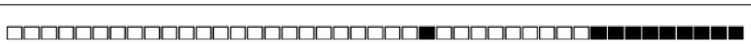
Lineage	Spoligotype Group	*n*	Spoligotyping pattern (spacers1-43)
1	EAI2	2	
	EAI2	1	
	EAI2	1	
2	Beijing	1	
	Beijing	1	
3	CAS	1	
	[a]CAS1Delhi	5	
4	T1	1	
	X1	2	
	H	1	
M.bovis	[b]X2	1	
M.bovis BCG	X	1	
Total		18	

Fig 3. Distribution of *in silico* generated spoligotypes across the culture-positive Tasmanian TB isolates analysed from 2014–2016. [a]The *in silico* derived spoligotype of the four Tasmanian Lineage 3 cluster isolates (RHH3, RHH11, RHH13, RHH14) and a fifth Lineage 3 isolate (RHH10) matched Spoligotype International Type 26 of the CAS1_Delhi spoligotyping family which accounted for approximately 50% of Lineage 3 *M. tuberculosis* isolates in Nepal in a previous analysis [21]. [b]The *in silico* derived spoligotype of the *M. bovis* isolate (TASMB14) matches that of human *M. bovis* cases that were reported in other Australian states/territories between 1977 and 1989 [22].

https://doi.org/10.1371/journal.pone.0192351.g003

First confirmed case of MDR-TB in Tasmania

An overseas-born individual tested positive for TB infection in an Interferon Gamma Release Assay (IGRA) test in 2016 but did not exhibit symptoms of TB, had a normal chest x-ray, and was sputum culture negative. The patient presented with developed abdominal pain consistent with colitis later in 2016 and a colon biopsy sample was subsequently taken. The colon tissue specimen was smear-negative but culture-positive for *Mycobacterium tuberculosis*. MGIT based drug-susceptibility testing performed on this extra-pulmonary isolate at VIDRL revealed that it was resistant to isoniazid, rifampicin, ethambutol, pyrazinamide making it the first confirmed case of MDR-TB to have occurred in Tasmania. The isolate was recorded as sensitive to ethionamide, amikacin, capreomycin, kanamycin, ofloxacin and moxifloxacin.

Genomic DNA of the Tasmanian MDR-TB isolate (TASMDR1) was sequenced on an Illumina MiSeq. Paired-end reads were mapped to the *M. tuberculosis* H37Rv reference genome by Burrows-Wheeler Alignment producing a mapped-read depth of 73.7-fold, covering 97.36% of the H37Rv genome. A consensus sequence was called using SAMtools generating a 4,320,496-bp draft assembly. With respect to reference H37Rv genome, 1,566 SLVs were detected in the TASMDR1 assembled genome, of which 874 were non-synonymous. An analysis was then performed to identify SLVs which correlated with phenotypic drug resistance. The genome of TASMDR1 displayed single-nucleotide polymorphisms in genes correlating

Table 2. Mutations detected in the genome of the TASMDR1 isolate that confer resistance to anti-tubercular drugs. Six mutations that have been associated with anti-tubercular drug resistance were identified. The mutations listed in the *rpoB*, *katG*, *pncA*, and *embB* genes were classified as high confidence SNPs with respect to resistance to rifampicin, isoniazid, pyrazinamide and ethambutol, respectively, by the PhyResSE database [12]. In addition, an A/C substitution was detected at position 514 of the 16S rRNA gene, *rrs* (MTB000019) that is associated with streptomycin resistance [33, 34].

Drug	Gene	Locus Tag	Mutation	Genome Location	Substitution
Rifampicin	*rpoB*	Rv0667	gAc / gGc, tCg /tTg	761110, 761155	D435G, S450L
Isoniazid	*katG*	Rv1908c	aGc / aCc	2155168	S315T
Pyrazinamide	*pncA*	Rv2043c	cCg / cTg	2289057	P62L
Ethambutol	*embB*	Rv3795	Atg / Gtg	4247429	M306V
Streptomycin	*rrs*	MTB000019	A / C	1472359	a514c

https://doi.org/10.1371/journal.pone.0192351.t002

with antimicrobial drug resistance when analysed using the PhyResSE database [12]. These included high confidence mutations in the genes *katG* (aGc/aCc, S315T) and *rpoB* (gAc/gGc, D435G; tCg/tTg, S450L) which are associated with *M. tuberculosis* resistance to isoniazid and rifampicin, respectively [26–28] (Table 2).

Further mutations were detected in the *embB* (Atg/Gtg, M306V) and *pncA* (cCg/cTg, P62L) genes that underlie resistance to ethambutol and pyrazinamide, respectively [29–32] (Table 2). In addition, an additional A/C substitution was detected at position 514 of the 16S rRNA gene, *rrs* (MTB000019) that is associated with streptomycin resistance [33, 34] (Table 2). The TASMDR1 isolate was predicted to belong to East Asian Lineage 2, sub-lineage Beijing, by the PhyResSE and TB Profiler databases [11, 12]. Furthermore, the isolate exhibited a polymorphism in the *mutT2* gene (Gga/Cga, G58R) which is associated with so-called 'Modern' Beijing strains [35, 36]. The patient was originally from Viet Nam and had known household contact with an active case of TB that was confirmed in Viet Nam in 2012. The isolate from this 2012 case was recorded as resistant to isoniazid, rifampicin, ethambutol, pyrazinamide and streptomycin from MGIT based drug-susceptibility testing.

Case of *M. bovis* TB in Tasmania

Isolate TASMB14 was collected in Tasmania in 2014 from a sputum specimen taken from a drug-susceptible case of pulmonary TB in an Australian-born person. Risk factors associated with this case included age (≥70 years) and chronic obstructive pulmonary disease co-morbidity. The genome sequence of the isolate revealed that it contains the Rv2043c (*pncA*) polymorphism, Cac/Gac, H57D, and the RD1 region genes Rv3871 to Rv3879c, confirming it as *M. bovis*. The *in silico* derived spoligotype of TASMB14 matches that of other human *M. bovis* cases that were reported elsewhere in Australia between 1977 and 1989 [22] (Fig 3).

Discussion

In this study, we provide the first in-depth analysis of the molecular epidemiology of tuberculosis in Tasmania. MTBC isolates collected from culture-positive cases of TB in Tasmania from 2014 to 2016, were examined. The most common lineage detected among the Tasmanian samples analysed was the East-African Indian Lineage 3 (33.3%) followed by the Euro-American Lineage 4 (22.2%), Indo-Oceanic Lineage 1 (22.2%), and the East-Asian Lineage 2 (11.1%) (Fig 1). Individual cases of TB due to *M. bovis* and *M. bovis* BCG were notified in 2014 and 2016, respectively.

Our whole-genome sequence analyses identified two possible clusters of *M. tuberculosis* among the Tasmanian cases, one belonging to Lineage 3 and the other belonging to Lineage 4. The Lineage 3 cluster, consisted of four isolates separated by zero SLVs which is indicative of recent transmission between the patients based on previously-established SNP thresholds [14,

15]. The isolates were collected within a three-month period from household contacts who originated from Nepal. The *in silico* generated spoligotype of the four Lineage 3 cluster isolates matches Spoligotype International Type 26 of the CAS1_Delhi spoligotyping family. This particular spoligotype was common among TB cases in Nepal, constituting approximately 50% of Lineage 3 isolates, and 20% of total TB isolates, in an earlier study (Fig 3) [21].

In this work, we describe the first documented case of MDR-TB in Tasmania. This case was detected in the second half of 2016 in an overseas-born individual who had earlier moved from Viet Nam to Tasmania. The isolate, TASMDR1, which belongs to the East-Asian Lineage 2, was confirmed as being resistant to isoniazid, rifampicin, ethambutol and pyrazinamide in phenotypic drug-susceptibility testing. Furthermore, genome sequencing identified an a514c mutation in the *rrs* locus (MTB000019) that is associated with streptomycin resistance. A household contact of the patient had been diagnosed with pulmonary MDR-TB in Viet Nam in 2012. The isolate from this 2012 case was recorded as resistant to isoniazid, rifampicin, ethambutol, pyrazinamide, and streptomycin in MGIT culture-based drug-susceptibility testing. Based on the equivalent drug-resistance profiles of the two MDR-TB cases, it is likely that the Tasmanian case contracted the MDR strain of *M. tuberculosis* from the household contact some time previously and that the infection remained latent until reactivating as extrapulmonary MDR-TB in 2016. A recent study by Fox and colleagues conducted in Viet Nam found that household contacts of patients with MDR-TB have a higher risk of becoming tuberculin-skin test positive and of developing active TB compared to contacts of drug-susceptible TB [37].

While the proportion of TB cases in Australia that are MDR is currently under 2% (22 MDR-TB cases out of 1,263 TB notifications in 2013), the estimated costs associated with treating a case of TB increase substantially when going from drug-susceptible TB (USD$17,000 in the USA, €10,282 in 15 EU countries, per case) to multi-drug resistant TB (USD$134,000 in the USA, €57,213 in 15 EU countries, per case) [38, 39]. It was previously estimated that management of one case of extensively drug-resistant (XDR) TB in 2012 cost Queensland Health in the region of AUD $500,000 [40]. Hence, vigilance will need to be maintained with respect to the tracing of contacts of previous TB cases, especially MDR-TB cases, and the early detection of drug resistance in Tasmanian isolates.

Human TB caused by *M. bovis* was reported in Tasmania in 2014, nearly 40 years after the last confirmed case of bovine TB in the state in 1975 [5]. The pulmonary form of disease was diagnosed in a male aged $\geq$70 years. The source of this infection is unknown but a possibility is reactivation of a latent *M. bovis* infection acquired during earlier rural exposure to *M. bovis* prior to the elimination of bovine TB disease in Tasmania. The *in silico* derived spoligotype of the isolate, TASMB14, matches that of previously-described human *M. bovis* cases that were reported in other Australian states and territories between 1977 and 1989 [22] (Fig 3). As noted in 1999 by Cousins et al., "because of the usual long incubation periods that can occur between infection and development of disease, and because of the possibility of disease reactivation, especially in elderly or immunocompromised patients, human tuberculosis caused by *M. bovis* is likely to continue to be diagnosed for many years to come" [41].

In the majority of the Tasmanian cases analysed from 2014 to 2016, 83.3% of patients (n = 15) were born overseas. This corresponds with 89.2% and 87.6% of TB notifications nationally recorded in the overseas-born population in 2012 and 2013, respectively [42]. A number of European studies have found that immigrants are not a major source of TB infection for the native-born population [43, 44]. Sandgren and colleagues in their systematic review concluded, that "TB in a foreign-born population does not have a significant influence on TB in the native population in EU/EEA" [45]. In our study, we did not find evidence of transmission of TB from the overseas-born cases to the Australian-born population. Nevertheless, targets have been set for low incidence jurisdictions by the World Health Organisation for

the pre-elimination of TB by 2035 (defined as <10 TB cases per million population), and the elimination of TB by 2050 (<1 TB case per million population) [46]. The incidence rate of TB in Tasmania currently sits at approximately 16 per million population (1.6/100,000) [42]. Hence, a 60% drop in TB cases by 2035, and a 95% drop in TB cases by 2050 are required in Tasmania for the state to meet international targets.

A common trend seen in low TB burden countries is decreasing TB in the native-born population and increasing TB in the migrant population as a proportion of total cases [47]. Therefore, bringing TB rates into line with the World Health Organisation's goals will require efforts to reduce TB incidence in the foreign-born population. A major emphasis in Australia is placed upon pre-immigration screening. Visa applicants who are 11 years or older must undergo a chest x-ray and potentially, other diagnostic tests. If active TB is found, Australian law does not permit the granting of a visa until the applicant has completed treatment and has been declared free of active TB [48]. Pre-entry screening of foreign-born individuals is not specifically designed for the detection of latent tuberculosis infection (LTBI). However, most TB cases among the foreign-born population in industrialised countries are believed to be due to reactivation of LTBI rather than continuation of an existing case of active TB [49, 50]. In addition, molecular epidemiological studies have found a strong association between the lineage of the MTBC strain isolated from a migrant patient and the predominant lineage found in their region of origin [23–25]. In our study, all Lineage 3 (Central Asian (CAS)/Delhi) cases were in individuals from the Indian sub-continent where this lineage is prevalent [51]. It is probable that a number of the overseas-born patients who presented with TB in Tasmania had acquired *M. tuberculosis* infection prior to their arrival in the state or in Australia. Therefore, further consideration will need to be given to the management of TB in the migrant population in Tasmania in order to reduce the incidence of the disease in the state.

Conclusions

In summary, our work provides the first extensive analysis of the molecular epidemiology of tuberculosis in Tasmania. It identified the presence of two phylogenetic clusters of identical isolates of *M. tuberculosis* which is indicative of recent transmission of TB among household contacts. In addition, it established the genetic basis of the resistance exhibited by Tasmania's first confirmed case of MDR-TB. Our study highlights that while the incidence of TB in Tasmania is comparatively low, challenges remain with regard to the management of the disease in the migrant population, particularly from high TB prevalence countries, which will need to be overcome for the state to meet the World Health Organisation's 2035 and 2050 TB eradication goals.

Author Contributions

Conceptualization: Ronan F. O'Toole.

Data curation: Janet A. Fyfe, Maria Globan.

Formal analysis: Micheál Mac Aogáin.

Funding acquisition: Louise A. Cooley, Ronan F. O'Toole.

Investigation: Louise A. Cooley, Greg Haug, Janet A. Fyfe, Maria Globan, Ronan F. O'Toole.

Methodology: Sanjay S. Gautam, Micheál Mac Aogáin.

Supervision: Ronan F. O'Toole.

Writing – original draft: Ronan F. O'Toole.

Writing – review & editing: Sanjay S. Gautam, Louise A. Cooley, Greg Haug, Ronan F. O'Toole.

References

1. Australian Bureau of Statistics. Census. 2011.
2. Roe M. Life over Death: Tasmanians and Tuberculosis: Tasmanian Historical Research Association; 1999.
3. Hall R. Notifiable diseases surveillance, 1917 to 1991. Communicable diseases intelligence. 1993; 17:226–36.
4. NNDSS Annual Report Working Group. Australia's notifiable disease status, 2014: Annual report of the National Notifiable Diseases Surveillance System. Communicable Diseases Intelligence: 2016.
5. More SJ, Radunz B, Glanville RJ. Lessons learned during the successful eradication of bovine tuberculosis from Australia. Vet Rec. 2015; 177(9):224–32. https://doi.org/10.1136/vr.103163 PMID: 26338937.
6. Department of Health and Human Services—Tasmanian Government. Notifiable Diseases 2014 Annual Report. 2016.
7. Stop TB Partnership. Mycobacteriology Laboratory Manual. 2014.
8. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12):1647–9. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367.
9. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998; 393(6685):537–44. Epub 1998/06/20. https://doi.org/10.1038/31159 PMID: 9634230.
10. Mac Aogain M, Gautam SS, Bower JE, Basu I, O'Toole RF. Draft Genome Sequence of a New Zealand Rangipo Strain of Mycobacterium tuberculosis. Genome Announc. 2016; 4(4). https://doi.org/10.1128/genomeA.00657-16 PMID: 27389273.
11. Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. 2015; 7 (1):51. https://doi.org/10.1186/s13073-015-0164-0 PMID: 26019726.
12. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol. 2015; 53(6):1908–14. https://doi.org/10.1128/JCM.00025-15 PMID: 25854485.
13. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52(5):696–704. PMID: 14530136.
14. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis. 2013; 13(2):137–46. https://doi.org/10.1016/S1473-3099(12)70277-3 PMID: 23158499.
15. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review. Tuberculosis (Edinb). 2016; 98:77–85. https://doi.org/10.1016/j.tube.2016.02.009 PMID: 27156621.
16. Hasnain SE, O'Toole RF, Grover S, Ehtesham NZ. Whole genome sequencing: a new paradigm in the surveillance and control of human tuberculosis. Tuberculosis (Edinb). 2015; 95(2):91–4. https://doi.org/10.1016/j.tube.2014.12.007 PMID: 25586521.
17. Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC Med. 2016; 14:21. https://doi.org/10.1186/s12916-016-0566-x PMID: 27005433.
18. Migliori GB, Lange C, Girardi E, Centis R, Besozzi G, Kliiman K, et al. Extensively drug-resistant tuberculosis is worse than multidrug-resistant tuberculosis: different methodology and settings, same results. Clin Infect Dis. 2008; 46(6):958–9. Epub 2008/02/22. https://doi.org/10.1086/528875 PMID: 18288911.
19. Denholm J, Coulter C, Bastian I, and the National Tuberculosis Advisory C. Defining a tuberculosis cluster or outbreak. Communicable diseases intelligence quarterly report. 2016; 40(3):E356–E9. PMID: 28278410.
20. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, Kato S, et al. TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. PLoS One. 2015; 10(11):e0142951. https://doi.org/10.1371/journal.pone.0142951 PMID: 26565975.

21. Malla B, Stucki D, Borrell S, Feldmann J, Maharjan B, Shrestha B, et al. First insights into the phylogenetic diversity of Mycobacterium tuberculosis in Nepal. PLoS One. 2012; 7(12):e52297. https://doi.org/10.1371/journal.pone.0052297 PMID: 23300635.

22. Cousins DV, Williams SN, Dawson DJ. Tuberculosis due to Mycobacterium bovis in the Australian population: DNA typing of isolates, 1970–1994. Int J Tuberc Lung Dis. 1999; 3(8):722–31. PMID: 10460106.

23. Yen S, Bower JE, Freeman JT, Basu I, O'Toole RF. Phylogenetic lineages of tuberculosis isolates in New Zealand and their association with patient demographics. Int J Tuberc Lung Dis. 2013; 17(7):892–7. Epub 2013/05/03. https://doi.org/10.5588/ijtld.12.0795 PMID: 23635796.

24. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2006; 103(8):2869–73. Epub 2006/02/16. https://doi.org/10.1073/pnas.0511240103 PMID: 16477032.

25. Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, et al. Major Mycobacterium tuberculosis lineages associate with patient country of origin. J Clin Microbiol. 2009; 47(4):1119–28. Epub 2009/02/14. https://doi.org/10.1128/JCM.02142-08 PMID: 19213699.

26. Rodwell TC, Valafar F, Douglas J, Qian L, Garfein RS, Chawla A, et al. Predicting Extensively Drug-Resistant Mycobacterium tuberculosis Phenotypes with Genetic Mutations. Journal of Clinical Microbiology. 2014; 52(3):781–9. https://doi.org/10.1128/JCM.02701-13 PMID: 24353002

27. Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, et al. Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. 2015; 6:7119. https://www.nature.com/articles/ncomms8119#supplementary-information PMID: 25960343

28. Siddiqi N, Shamim M, Hussain S, Choudhary RK, Ahmed N, Prachee, et al. Molecular characterization of multidrug-resistant isolates of Mycobacterium tuberculosis from patients in North India. Antimicrob Agents Chemother. 2002; 46(2):443–50. https://doi.org/10.1128/AAC.46.2.443-450.2002 PMID: 11796356.

29. Starks AM, Gumusboga A, Plikaytis BB, Shinnick TM, Posey JE. Mutations at embB Codon 306 Are an Important Molecular Indicator of Ethambutol Resistance in Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy. 2009; 53(3):1061–6. https://doi.org/10.1128/AAC.01357-08 PMID: 19104018

30. Cuevas-Córdoba B, Juárez-Eusebio DM, Almaraz-Velasco R, Muñiz-Salazar R, Laniado-Laborin R, Zenteno-Cuevas R. Mutation at embB Codon 306, a Potential Marker for the Identification of Multidrug Resistance Associated with Ethambutol in Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy. 2015; 59(9):5455–62. https://doi.org/10.1128/AAC.00117-15 PMID: 26124153

31. Miotto P, Cabibbe AM, Feuerriegel S, Casali N, Drobniewski F, Rodionova Y, et al. Mycobacterium tuberculosis Pyrazinamide Resistance Determinants: a Multicenter Study. mBio. 2014; 5(5). https://doi.org/10.1128/mBio.01819-14 PMID: 25336456

32. Simons SO, van der Laan T, Mulder A, van Ingen J, Rigouts L, Dekhuijzen PNR, et al. Rapid diagnosis of pyrazinamide-resistant multidrug-resistant tuberculosis using a molecular-based diagnostic algorithm. Clinical Microbiology and Infection. 2014; 20(10):1015–20. https://doi.org/10.1111/1469-0691.12696 PMID: 24890253

33. Ballif M, Harino P, Ley S, Coscolla M, Niemann S, Carter R, et al. Drug resistance-conferring mutations in Mycobacterium tuberculosis from Madang, Papua New Guinea. BMC Microbiology. 2012; 12:191-. https://doi.org/10.1186/1471-2180-12-191 PMID: 22943573

34. Ritter C, Lucke K, Sirgel FA, Warren RW, van Helden PD, Böttger EC, et al. Evaluation of the AID TB Resistance Line Probe Assay for Rapid Detection of Genetic Alterations Associated with Drug Resistance in Mycobacterium tuberculosis Strains. Journal of Clinical Microbiology. 2014; 52(3):940–6. https://doi.org/10.1128/JCM.02597-13 PMID: 24403306

35. Yang C, Luo T, Sun G, Qiao K, Sun G, DeRiemer K, et al. Mycobacterium tuberculosis Beijing Strains Favor Transmission but Not Drug Resistance in China. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America. 2012; 55(9):1179–87. https://doi.org/10.1093/cid/cis670 PMID: 22865872

36. Mestre O, Luo T, Dos Vultos T, Kremer K, Murray A, Namouchi A, et al. Phylogeny of Mycobacterium tuberculosis Beijing Strains Constructed from Polymorphisms in Genes Involved in DNA Replication, Recombination and Repair. PLOS ONE. 2011; 6(1):e16020. https://doi.org/10.1371/journal.pone.0016020 PMID: 21283803

37. Fox GJ, Anh NT, Nhung NV, Loi NT, Hoa NB, Ngoc Anh LT, et al. Latent tuberculous infection in household contacts of multidrug-resistant and newly diagnosed tuberculosis. The International Journal of Tuberculosis and Lung Disease. 2017; 21(3):297–302. https://doi.org/10.5588/ijtld.16.0576 PMID: 28225339

38. Diel R, Vandeputte J, de Vries G, Stillo J, Wanlin M, Nienhaus A. Costs of tuberculosis disease in the European Union: a systematic analysis and cost calculation. European Respiratory Journal. 2014; 43 (2):554–65. https://doi.org/10.1183/09031936.00079413 PMID: 23949960

39. Bhakta S, Besra GS, Upton AM, Parish T, Sholto-Douglas-Vernon C, Gibson KJ, et al. Arylamine N-acetyltransferase is required for synthesis of mycolic acids and complex lipids in Mycobacterium bovis BCG and represents a novel drug target. J Exp Med. 2004; 199(9):1191–9. Epub 2004/05/01. https://doi.org/10.1084/jem.20031956 PMID: 15117974.

40. Kirby T. Extensively drug-resistant tuberculosis hovers threateningly at Australia's door. Med J Aust. 2013;(198 (7): 355). PMID: 23581944

41. Cousins DV, Dawson DJ. Tuberculosis due to Mycobacterium bovis in the Australian population: cases recorded during 1970–1994. Int J Tuberc Lung Dis. 1999; 3(8):715–21. PMID: 10460105.

42. Toms C, Stapledon R, Waring J, Douglas P. Tuberculosis notifications in Australia, 2012 and 2013. Communicable diseases intelligence quarterly report. 2015; 39(2):E217–35. Epub 2015/08/04. PMID: 26234258.

43. Barniol J, Niemann S, Louis VR, Brodhun B, Dreweck C, Richter E, et al. Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant sub-populations. BMC Infectious Diseases. 2009; 9(1):197. https://doi.org/10.1186/1471-2334-9-197 PMID: 19961606

44. Kamper-Jørgensen Z, Andersen AB, Kok-Jensen A, Kamper-Jørgensen M, Bygbjerg IC, Andersen PH, et al. Migrant tuberculosis: the extent of transmission in a low burden country. BMC Infectious Diseases. 2012; 12(1):60. https://doi.org/10.1186/1471-2334-12-60 PMID: 22423983

45. Sandgren A, Sañé Schepisi M, Sotgiu G, Huitric E, Migliori GB, Manissero D, et al. Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: a systematic review. European Respiratory Journal. 2014; 43(4):1159–71. https://doi.org/10.1183/09031936.00117213 PMID: 24114966

46. Parish T, Turner J, Stoker NG. amiA is a negative regulator of acetamidase expression in Mycobacterium smegmatis. BMC Microbiol. 2001; 1:19. Epub 2001/09/26. PMID: 11570974.

47. Hanway A, Comiskey CM, Tobin K, O'Toole RF. Relating annual migration from high tuberculosis burden country of origin to changes in foreign-born tuberculosis notification rates in low-medium incidence European countries. Tuberculosis. 101:67–74. https://doi.org/10.1016/j.tube.2016.07.019 PMID: 27865401

48. Department of Immigration and Border Protection—Australian Government. Fact Sheet—Health Requirement. 2015.

49. White RG, Houben RMGJ. Towards elimination in industrialised countries: expanding diagnosis and treatment of LTBI among immigrants [Editorial]. The International Journal of Tuberculosis and Lung Disease. 2014; 18(4):380-. https://doi.org/10.5588/ijtld.14.0156 PMID: 24670689

50. Ricks PM, Cain KP, Oeltmann JE, Kammerer JS, Moonan PK. Estimating the Burden of Tuberculosis among Foreign-Born Persons Acquired Prior to Entering the U.S., 2005–2009. PLOS ONE. 2011; 6 (11):e27405. https://doi.org/10.1371/journal.pone.0027405 PMID: 22140439

51. Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, et al. Predominance of ancestral lineages of Mycobacterium tuberculosis in India. Emerg Infect Dis. 2006; 12(9):1367–74. Epub 2006/11/01. https://doi.org/10.3201/eid1209.050017 PMID: 17073085.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Draft Genome Sequence of the First Confirmed Isolate of Multidrug-Resistant *Mycobacterium tuberculosis* in Tasmania

Sanjay S. Gautam,[a] Micheál Mac Aogáin,[b] Ronan F. O'Toole[a,b]

School of Medicine, Faculty of Health, University of Tasmania, Hobart, Tasmania, Australia[a]; Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, St. James's Hospital, Dublin, Ireland[b]

ABSTRACT The spread of multidrug-resistant (MDR) tuberculosis (TB) has become a major global challenge. In 2016, Tasmania recorded its first known incidence of MDR-TB. Here, we report the draft whole-genome sequence of the *Mycobacterium tuberculosis* isolate from this case, TASMDR1, and describe single-nucleotide polymorphisms associated with its drug resistance.

Received 29 September 2017 **Accepted** 2 October 2017 **Published** 2 November 2017

Citation Gautam SS, Mac Aogáin M, O'Toole RF. 2017. Draft genome sequence of the first confirmed isolate of multidrug-resistant *Mycobacterium tuberculosis* in Tasmania. Genome Announc 5:e01230-17. https://doi.org/10.1128/genomeA.01230-17.

Address correspondence to Ronan F. O'Toole, ronan.otoole@utas.edu.au.

The earliest written record of tuberculosis (TB) in Tasmania comes from Colonel David Collins, who reported in 1804 that a member of his Hobart settlement had consumption (1). Today, Tasmania is regarded as a low-TB-burden state with an incidence rate of 1.7/100,000 persons compared to 5.7/100,000 nationally in Australia in 2014 (2). Until recently, Tasmania had been free of multidrug-resistant (MDR) forms of TB; however, in 2016 its first case of MDR-TB was reported.

We have previously characterized the genomes of MDR and extensively drug-resistant (XDR) *Mycobacterium tuberculosis* (3–5). Here, genomic DNA of the Tasmanian isolate, TASMDR1, was sequenced using an Illumina MiSeq instrument. A total of 2,860,297 paired-end reads were mapped to the publicly available annotated genome of *M. tuberculosis* reference strain H37Rv (GenBank accession number NC_000962.3) (6) by Burrows-Wheeler alignment (7). This yielded an average read depth of 65.5-fold, covering 97.8% of the H37Rv genome. Variants relative to the H37Rv reference genome were called using the SAMtools analysis suite, and variant annotation was performed using SnpEff (8, 9). A 4,230,496-bp draft genome assembly of 220 contigs was assembled *de novo* using the SPAdes assembler (v3.7) (10). Assembled contigs were ordered with respect to the *M. tuberculosis* H37Rv genome using ABACAS (11).

A total of 1,553 variant sites were identified relative to the H37Rv genome and consisted of 1,408 single-nucleotide variants (SNVs) and 145 insertions/deletions. Of the variants, 881 were nonsynonymous; of these, 784 were SNVs and 97 were insertions/deletions. The genome of TASMDR1 displayed high-confidence single-nucleotide polymorphisms in genes correlating with antimicrobial drug resistance when analyzed using the PhyResSE database (12). These include high-confidence mutations in the *katG* gene (aGc/aCc, S315T) and *rpoB* gene (gAc/gGc, D435G; tCg/tTg, S450L), which underlie *M. tuberculosis* resistance to isoniazid and rifampin, respectively (13, 14). These data establish the genetic bases of the MDR phenotype exhibited by strain TASMDR1.

Additional mutations were detected in the *embB* gene (Atg/Gtg, M306V) and *pncA* gene (cCg/cTg, P62L) that are associated with resistance to ethambutol and pyrazinamide, respectively (15–18). Furthermore, an A/C substitution was detected at position 514 of the 16S rRNA gene, *rrs* (MTB000019), and is related to streptomycin resistance (19, 20). The TASMDR1 isolate belongs to the Beijing sublineage of East Asian Lineage 2, as predicted by the PhyResSE and TB Profiler databases (12, 21).

The drug-resistance mutations that were identified in the genome of the TASMDR1 isolate were detected within a significantly shorter turn-around time compared to

conventional phenotypic drug-susceptibility testing. Although MDR-TB isolates are currently rare in Tasmania, this study highlights the utility of having a microbial whole-genome sequencing facility available for rapidly determining the resistance profiles of MDR-TB isolates that may present in a low-TB-incidence setting.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number NTFG00000000. The version described in this paper is version NTFG01000000.

ACKNOWLEDGMENTS

R.F.O. was the recipient of a Royal Hobart Hospital Research Foundation grant (17-104). S.S.G. was the recipient of a School of Medicine/Faculty of Health Ph.D. scholarship. M.M.A. was the recipient of an Irish Research Council fellowship (EPSPD/2015/32).

REFERENCES

1. Roe M. 1999. Life over death: Tasmanians and tuberculosis: Hobart, Tasmania: Tasmanian Historical Research Association.
2. NNDSS Annual Report Working Group. 2016. Australia's notifiable disease status, 2014: annual report of the National Notifiable Diseases Surveillance System. *Commun Dis Intell* 40:E48–E145.
3. O'Toole RF, Johari BM, Mac Aogáin M, Rogers TR, Bower JE, Basu I, Freeman JT. 2014. Draft genome sequence of the first isolate of extensively drug-resistant *Mycobacterium tuberculosis* in New Zealand. Genome Announc 2(3):1–2. https://doi.org/10.1128/genomeA.00319-14.
4. Roycroft E, Mac Aogáin M, O'Toole RF, Fitzgibbon M, Rogers TR. 2014. Draft genome sequence of the first isolate of extensively drug-resistant *Mycobacterium tuberculosis* in Ireland. Genome Announc 2(5):e01002-14. https://doi.org/10.1128/genomeA.01002-14.
5. Mac Aogáin M, Johari BM, Bower JE, O'Toole RF. 2014. Draft genome sequence of a multidrug-resistant New Zealand isolate of *Mycobacterium tuberculosis* lineage 3. Genome Announc 2(5):e01017-14. https://doi.org/10.1128/genomeA.01017-14.
6. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544. https://doi.org/10.1038/31159.
7. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
8. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:80–92. https://doi.org/10.4161/fly.19695.
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
11. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25:1968–1969. https://doi.org/10.1093/bioinformatics/btp347.
12. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, Cabibbe AM, Niemann S, Fellenberg K. 2015. PhyResSE: a Web Tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. J Clin Microbiol 53:1908–1914. https://doi.org/10.1128/JCM.00025-15.
13. Rodwell TC, Valafar F, Douglas J, Qian L, Garfein RS, Chawla A, Torres J, Zadorozhny V, Kim MS, Hoshide M, Catanzaro D, Jackson L, Lin G, Desmond E, Rodrigues C, Eisenach K, Victor TC, Ismail N, Crudu V, Gler MT, Catanzaro A. 2014. Predicting extensively drug-resistant *Mycobacterium tuberculosis* phenotypes with genetic mutations. J Clin Microbiol 52:781–789. https://doi.org/10.1128/JCM.02701-13.
14. Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, Balloux F. 2015. Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. Nat Commun 6:7119. https://doi.org/10.1038/ncomms8119.
15. Starks AM, Gumusboga A, Plikaytis BB, Shinnick TM, Posey JE. 2009. Mutations at embB codon 306 are an important molecular indicator of ethambutol resistance in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 53:1061–1066. https://doi.org/10.1128/AAC.01357-08.
16. Cuevas-Córdoba B, Juárez-Eusebio DM, Almaraz-Velasco R, Muñiz-Salazar R, Laniado-Laborin R, Zenteno-Cuevas R. 2015. Mutation at embB codon 306, a potential marker for the identification of multidrug resistance associated with ethambutol in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 59:5455–5462. https://doi.org/10.1128/AAC.00117-15.
17. Miotto P, Cabibbe AM, Feuerriegel S, Casali N, Drobniewski F, Rodionova Y, Bakonyte D, Stakenas P, Pimkina E, Augustynowicz-Kopeć E, Degano M, Ambrosi A, Hoffner S, Mansjö M, Werngren J, Rüsch-Gerdes S, Niemann S, Cirillo DM. 2014. *Mycobacterium tuberculosis* pyrazinamide resistance determinants: a multicenter study. mBio 5:e01819-14. https://doi.org/10.1128/mBio.01819-14.
18. Simons SO, van der Laan T, Mulder A, van Ingen J, Rigouts L, Dekhuijzen PNR, Boeree MJ, van Soolingen D. 2014. Rapid diagnosis of pyrazinamide-resistant multidrug-resistant tuberculosis using a molecular-based diagnostic algorithm. Clin Microbiol Infect 20:1015–1020. https://doi.org/10.1111/1469-0691.12696.
19. Ballif M, Harino P, Ley S, Coscolla M, Niemann S, Carter R, Coulter C, Borrell S, Siba P, Phuanukoonnon S, Gagneux S, Beck HP. 2012. Drug resistance-conferring mutations in *Mycobacterium tuberculosis* from Madang, Papua New Guinea. BMC Microbiol 12:191–191. https://doi.org/10.1186/1471-2180-12-191.
20. Ritter C, Lucke K, Sirgel FA, Warren RW, van Helden PD, Böttger EC, Bloemberg GV. 2014. Evaluation of the AID TB resistance line probe assay for rapid detection of genetic alterations associated with drug resistance in *Mycobacterium tuberculosis* strains. J Clin Microbiol 52:940–946. https://doi.org/10.1128/JCM.02597-13.
21. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 7:51. https://doi.org/10.1186/s13073-015-0164-0.

Intercontinental translocation of latent multidrug-resistant tuberculosis to Australia demonstrated by whole genome sequencing

To the Editor: In 2016, there were an estimated 490 000 cases globally of multidrug-resistant (MDR) tuberculosis exhibiting resistance to isoniazid and rifampicin.[1] The first case of MDR tuberculosis diagnosed in Tasmania occurred in 2016 in a Vietnamese-born person. Vietnam was the second highest reported country of birth for overseas-born patients with tuberculosis notified in Australia in 2014.[2] The patient had previously tested positive for tuberculosis infection in an interferon-γ release assay test performed in Tasmania in early 2016, but at the time, the patient was asymptomatic and had a normal chest x-ray and a negative sputum culture. After an episode of colitis, a colon tissue biopsy specimen isolated *Mycobacterium tuberculosis*. Whole genome sequence of the isolate (TASMDR1), identified high confidence mutations for isoniazid, rifampicin, ethambutol and pyrazinamide, in accordance with the culture-based drug susceptibility testing, and, in addition, it identified a mutation associated with streptomycin resistance.[3]

We became aware that a household contact of the Tasmania-located patient with MDR tuberculosis had been diagnosed with pulmonary tuberculosis in Vietnam in 2012 and requested the drug susceptibility testing data for this isolate (VTB1) from the treating hospital in Ho Chi Minh City. VTB1 was resistant to isoniazid, rifampicin, ethambutol, pyrazinamide and streptomycin in culture-based drug susceptibility testing. We therefore obtained a genomic DNA preparation of VTB1 to enable direct comparison with the TASMDR1 isolate collected in Tasmania. Next generation sequencing of VTB1 was performed on an Illumina platform and paired-end reads were mapped to the *M. tuberculosis* H37Rv reference genome (NC_000962.3). The Box shows variants associated with drug resistance.

In addition to drug resistance mutations, VTB1 shared all previously described variants in TASMDR1 with respect to H37Rv[3] and, therefore, the two isolates were genetically indistinguishable. This is strongly indicative of transmission involving the two patients based on established single nucleotide polymorphism thresholds.[4] It is most probable that the patient diagnosed in Tasmania contracted the MDR strain of *M. tuberculosis* during the episode of pulmonary disease diagnosed in the household contact in 2012 and that the infection remained latent until reactivating as extrapulmonary MDR tuberculosis in 2016.

In conclusion, the global burden of latent tuberculosis infection has been estimated to be 23% of the world's population, which corresponds to about 1.7 billion people.[5] Despite the immense prevalence of latent tuberculosis infection, there are few reports in the literature that confirm using genome variant analyses for the translocation of the MDR form of tuberculosis from one jurisdiction to another as a latent infection and its subsequent emergence as active MDR tuberculosis in a new host country. This type of transit of tuberculosis is difficult to detect with pre-immigration screening practices that are reliant upon a diagnosis of pulmonary tuberculosis based on a chest x-ray. The international movement of MDR tuberculosis in latent form, as has been determined in this case, is an area of concern and could be a significant challenge for future tuberculosis eradication. The growing application of genome sequencing in tuberculosis diagnostics and surveillance will help establish the level of MDR tuberculosis cases due to reactivation of latent tuberculosis infection.

Sanjay S Gautam[1]
Greg Haug[2]
Louise A Cooley[3]
Micheál Mac Aogáin[4]
Ronan F O'Toole[1,4,5]

1 University of Tasmania, Hobart, TAS.
2 Launceston General Hospital, Launceston, TAS.
3 Royal Hobart Hospital, Hobart, TAS.
4 Trinity College Dublin, University of Dublin, Dublin, Ireland.
5 La Trobe University, Wodonga, VIC.

Ronan.OToole@utas.edu.au

Acknowledgements: This research was supported by funding from the Royal Hobart Hospital Research Foundation (grant reference no. 17-104). The ethics approval for this study was obtained from the Tasmanian Health and Medical Human Research Ethics Committee (approval no. H0016214).

Competing interests: No relevant disclosures. ■

doi: 10.5694/mja2.50044

References are available online.

Drug resistance determining mutations in a contact who presented with tuberculosis in Vietnam in 2012 (VTB1) and in the first confirmed patient with multidrug-resistant (MDR) tuberculosis in Tasmania in 2016 (TASMDR1). The isolate from VTB1 and TASMDR1 share identical drug resistance mutations

Isolates	Drug	Gene	Function	Mutation	Substitution
VTB1 and TASMDR1	Rifampicin	*rpoB (Rv0667)*	RNA polymerase β-subunit	gAc/gGc, tCg/tTg	D435G, S450L
	Isoniazid	*katG (Rv1908c)*	Catalase-peroxidase	aGc/aCc	S315T
	Pyrazinamide	*pncA (Rv2043c)*	Pyrazinamidase/ nicotinamidase	cCg/cTg	P62L
	Ethambutol	*embB (Rv3795)*	Arabinosyltransferase B	Atg/Gtg	M306V
	Streptomycin	*rrs (MTB000019)*	16S ribosomal RNA	a/c	a514c*

* Substitution located in a non-protein coding gene. ◆

1 World Health Organization. Global tuberculosis report, 2017. https://www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf (viewed Feb 2019).
2 Toms C, Stapledon R, Coulter C, Douglas P. Tuberculosis notifications in Australia, 2014. *Commun Dis Intell Q Rep* 2017; 41: E247–E263.
3 Gautam SS, Mac Aogáin M, Cooley LA, et al. Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. *PLoS ONE* 2018; 13: e0192351.
4 Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013; 13: 137–146.
5 Houben RM, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS Med* 2016; 13: e1002152. ■

Journal of Biological Methods | 2019 | Vol. 6(1) | e110
DOI: 10.14440/jbm.2019.276

PROTOCOL

A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens

Sanjay S. Gautam[1], Rajendra KC[1], Kelvin WC Leong[3], Micheál Mac Aogáin[2], Ronan F. O'Toole[1,2,3]*

[1]School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, TAS 7001, Australia
[2]Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, Dublin D08 W9RT, Ireland
[3]School of Molecular Sciences, College of Science, Health and Engineering, La Trobe University, Victoria 3690, Australia

*Corresponding Author: Ronan F. O'Toole, Email: r.otoole@latrobe.edu.au

Competing interests: The authors have declared that no competing interests exist.

Abbreviations used: DAL, diluted amplified library; NGS, next-generation sequencing; NTA, Nextera XT tagment amplicon; PAL, pooled amplified library; PCR, polymerase chain reaction; VNTR, variable number tandem repeat; WES, whole exome sequencing; WGS, whole genome sequencing

Received September 28, 2018; Revision received January 13, 2019; Accepted January 28, 2019; Published March 15, 2019

ABSTRACT

Bacterial whole genome sequencing (WGS) is becoming a widely-used technique in research, clinical diagnostic, and public health laboratories. It enables high resolution characterization of bacterial pathogens in terms of properties that include antibiotic resistance, molecular epidemiology, and virulence. The introduction of next-generation sequencing instrumentation has made WGS attainable in terms of costs. However, the lack of a beginner's protocol for WGS still represents a barrier to its adoption in some settings. Here, we present detailed step-by-step methods for obtaining WGS data from a range of different bacteria (Gram-positive, Gram-negative, and acid-fast) using the Illumina platform. Modifications have been performed with respect to DNA extraction and library normalization to maximize the output from the laboratory consumables invested. The protocol represents a simplified and reproducible method for producing high quality sequencing data. The key advantages of this protocol include: simplicity of the protocol for users with no prior genome sequencing experience and reproducibility of the protocol across a wide range of bacteria.

Keywords: whole genome sequencing, *Enterococcus faecium*, *Haemophilus influenzae*, *Mycobacterium tuberculosis*

BACKGROUND

Using Sanger sequencing, the Human Genome Project expended approximately USD $2.7 billion and took more than 10 years to produce the first human genome sequence. Today, a human genome can be sequenced in a matter of days for less than USD $1000 on a single next-generation sequencing (NGS) machine. This step change in throughput and per-base cost has transformed the use of DNA sequencing in biomedical research and is being translated in an expanding number of ways into medicine. NGS is increasingly being applied to understanding and managing infectious diseases. This includes the sequencing of microbial genomes for the purposes of laboratory identification of infectious agents [1], detection of antibiotic resistance markers [2], and the public health surveillance of epidemiological clusters and outbreaks [3]. Examples include its deployment in public health surveillance and control of community cases of *Escherichia coli* [4], *Campylobacter jejuni* [5], *Legionella pneumophila* [6] and *Mycobacterium tuberculosis* [7] disease, or global and regional epidemics caused by influenza [8], Ebola [9], and Zika [10] viruses. It has also been utilised to track the source and spread of healthcare-associated infections caused by *Staphylococcus aureus* [11], *Pseudomonas aeruginosa* [12], *Acinetobacter baumannii* [13], and *Enterococcus faecium* [14] in order to guide infection prevention and control in hospitals.

In addition to its whole genome (WGS), whole exome (WES), transcriptome (RNA-Seq), bisulphite methylome, and metagenomic sequencing capabilities, NGS can be directed to the detection of specific genes or mutations associated with human disease through targeted-panel amplicon screening. However, barriers remain with regard to establishing NGS in a laboratory for the first time and this hinders its uptake in clinical microbiology and other settings. One of these challenges is the lack of a simplified step-by-step protocol that can be picked up by laboratory personnel with no prior training or experience in NGS and used to generate reliable, high quality sequence data. Illumina dye-sequencing is currently considered the gold standard internationally in terms of read depth and base-calling accuracy, genome coverage, scalability, and the range of sequencing applications it delivers.

In this work, we produced an easy-to-follow, step-by-step NGS protocol with consistent genome coverage and average read depth that

How to cite this article: Gautam SS, Rajendra KC, Leong KWC, Mac Aogáin M, O'Toole RF. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. *J Biol Methods* 2019;6(1):e110. DOI: 10.14440/jbm.2019.276

was applicable to a range of bacterial pathogens *i.e.*, Gram-positive vancomycin-resistant *Enterococcus faecium*, Gram-negative non-typeable *Haemophilus influenzae*, and acid-fast high-GC content *Mycobacterium tuberculosis*. This protocol can be used to generate Illumina-based WGS data for clinical isolates of bacterial pathogens of importance to human health.

Figure 1 is the graphical summary of the process of obtaining whole genome sequence data from bacterial culture. This wet laboratory procedure generated FastQ reads from the sequencer within three days of start. We modified a number of the DNA extraction steps to obtain a sufficient quantity of contamination free template. Similarly, we replaced library normalization plates and Nextera XT tagment amplicon (NTA) plates with conventional polymerase chain reaction (PCR) tubes which may represent a cost-effective alternative. In addition, we have recommended the use of equal DNA concentrations of each library during library normalization to ensure better coverage and minimize bias. Simplification of bacterial NGS may assist in its uptake by beginner users.

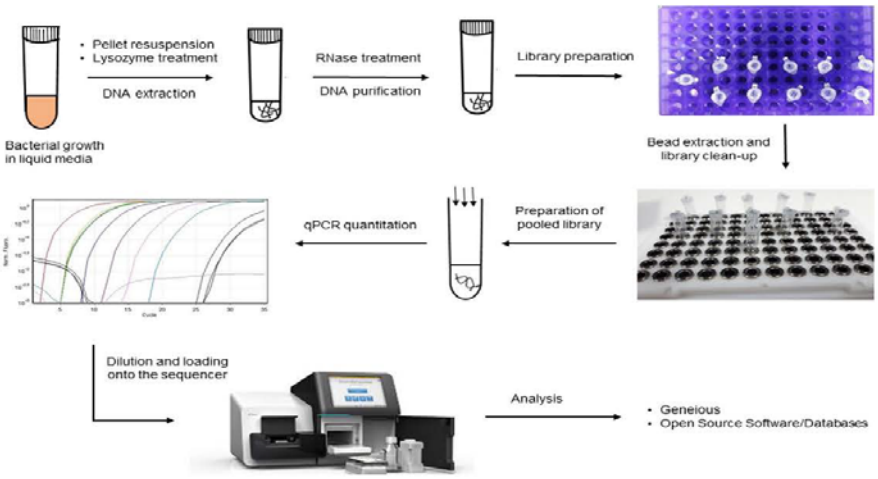


Figure 1. Graphical summary of the process of obtaining whole genome sequence data from a bacterial culture.

MATERIALS

Reagents

- ✓ Lysozyme (VWR, Australia, Cat.# 0663-10G)
- ✓ Ethanol, Pure (Sigma-Aldrich, Australia, Cat. # E7023)
- ✓ 2-Propanol (Sigma-Aldrich, Australia, Cat. # I8912)
- ✓ Phosphate Buffered Saline (Gibco™- Thermo Fisher Scientific, UK, Cat. # 10010023)
- ✓ Ultrapure™ DNase/RNase Free Distilled Water (Invitrogen, Australia Cat. # 10977-015)
- ✓ DNeasy® Blood and Tissue Kit (Qiagen, Germany, Cat. # 69504)
- ✓ High Pure PCR Template Preparation Kit (Roche, Germany, Cat. # 11796828001)
- ✓ Qubit™ dsDNA HS (High Sensitivity) Assay Kit (Invitrogen, Australia, Cat. # Q32851)
- ✓ Nextera® DNA Library Preparation Kit (Illumina, USA, Cat. # FC-121-1030)
- ✓ Nextera® XT Library Preparation Kit (Illumina, USA, Cat. # FC-131-1024)
- ✓ Nextera® XT Index Kit (Illumina, USA, Cat. # FC-131-1001)
- ✓ Miseq Reagent Kit v2 (300 cycles) (Illumina, USA, Cat. # MS-102-2002)
- ✓ KAPA Library Quantification Kit (Illumina, USA, Cat. # 07960140001)
- ✓ Agencourt® AMPure XP beads (Beckman Coulter, USA, Cat. # A63880)

Recipes

- ✓ Qubit working solution: dilute Qubit dsDNA HS Reagent 1:200 in Qubit dsDNA HS buffer. For *n* samples, prepare *n* × 200 µl working solution.
- ✓ 80% ethanol: add 2 ml absolute ethanol into 8 ml distilled water.
- ✓ 0.2 M NaOH: weigh 0.04 g of NaOH pellet and dissolve it into 5 ml distilled water.

Equipment

- ✓ Qubit™ assay tubes (Life-technologies, USA, Cat. # Q32856)
- ✓ PCR tubes (Molecular Bioproducts, USA, Cat. # MBP3412)
- ✓ Qubit® 2.0 Fluorometer (Invitrogen, Australia, Cat. # Q32866)
- ✓ Agencourt Magnetic stand (Beckman Coulter, USA, Cat. # A32782)
- ✓ Applied Biosystems® Veriti 96-Well thermal cycler (Thermo Fisher Scientific, USA)
- ✓ Rotor-Gene 6000 real-time thermocycler (Corbett Research, Australia)

PROCEDURE

Extraction of bacterial genomic DNA

1. Pellet the liquid culture media (200 µl) by centrifuging at 8000 *g* for 8 min in a sterile microfuge tube.

   CRITICAL STEP: All bacterial cultures should be treated as potentially pathogenic to the laboratory worker and colleagues. Therefore, the use of appropriate aseptic techniques, and the wearing of appropriate personal protective equipment are strongly recommended to maintain acceptable work health and safety standards and minimise exposure to harmful agents.

2. Resuspend the pellet in 600 µl phosphate-buffered saline (1×) until the absorbance at 600 nm (A_{600}) is between 1.0 and 2.0. Lyse the cells by adding 30 µl lysozyme (50 mg/ml), vortex, and incubate at 37°C for 1 h.
3. Follow the DNeasy® Blood and Tissue Kit Quick-start protocol to extract the DNA.
4. Elute the DNA in 100 µl volume and treat it with 2 µl RNase (100 mg/ml) (Qiagen, Hilden, Germany) and incubate at room temperature for 1 h.
5. Purify RNase-treated DNA using the High Pure PCR Template Preparation Kit.

   TIP: Perform only 4 DNA spin-wash steps instead of 9 recommended steps. Pre-incubate the elution buffer in a heat block set at 70°C.

6. Add 100 µl of binding buffer to RNase treated DNA and incubate at 70°C for 10 min.
7. Add 50 µl of 2-Propanol and transfer the content to a Roche spin column and spin at 8000 *g* for 1 min.
8. Discard the flow through and insert the spin column into a new collection tube.
9. Wash by adding 500 µl wash buffer and spin at 8000 *g* for 1 min.
10. Discard the flow through and insert the spin column into a new collection tube.
11. Perform a final spin at 8000 *g* for 1 min.
12. Finally, insert the column into a 1.5 ml sterile microfuge tube, add 50 µl of pre-heated elution buffer and spin at 8000 *g* for 1 min to elute the purified DNA for next generation sequencing.

    CRITICAL STEP: For next generation sequencing, contaminant-free, high-molecular weight DNA with an absorbance (260 nm/280 nm) ratio between 1.8 to 2.0 is considered a high-quality template DNA [15].

Quantification of bacterial genomic DNA

13. Dispense 190 µl and 198 µl of Qubit working solution in standard and sample tubes, respectively.
14. Add 10 µl standards (1 and 2) and 2 µl of sample in separate Qubit assay tubes.
15. Vortex the mixture for 3 s and incubate at room temperature for 2 min before taking the reading.
16. Adjust the DNA concentration of each sample to 0.2 ng/µl by diluting with a required volume of distilled water.

    CRITICAL STEP: The use of an accurate concentration of DNA is crucial for bacterial DNA genomic library preparation.

Tagmentation and PCR amplification of bacterial genomic DNA

TIP: For all of the methods below, the recommended 96-well TYC plate can be replaced with 0.2 ml thin wall clear, flat capped PCR tubes. In addition, multichannel pipettes and the high-speed micro plate shaker can be replaced with single channel pipettes and a bench top centrifuge, respectively.

Nextera XT tagment amplicon construction

17. In a PCR tube, add 5 µl tagmentation DNA buffer and 2.5 µl amplification tagmentation mix to 2.5 µl (0.2 ng/µl) input DNA.

18. Briefly vortexed the content and transfer to the thermocycler programmed for one step at 55°C for 5 min with heated lid, followed by a hold at 10°C for a volume of 10 µl.

Neutralization of Nextera XT tagment amplicon

19. Immediately after reaching the hold temperature of 10°C in the above step, neutralize NTA by adding 2.5 µl neutralization tagmentation buffer and incubate at room temperature for 5 min.

PCR amplification

20. For amplification, add 7.5 µl Nextera® PCR mastermix and 2.5 µl of each index primer, 1 and 2, to a tube containing neutralized NTA.

    CRITICAL STEP: Primer combinations, S502 with N705 /706 and S503 with N701/702 should be avoided. Avoid any repeated combinations and carefully note the primers used for each sample.

21. Gently pipette the content and perform a quick spin.
22. Proceed to amplification in a thermocycler programmed for a working volume of 25 µl with the following settings: heated lid, initial cycle at 72°C for 3 min followed by 95°C for 30 s and 12 cycles of (95°C for 10 s, 55°C for 30 s and 72°C for 30 s) with a final run at 72°C for 5 min followed by a hold temperature of 10°C. The amplified, tagmented library can be stored at 2 to 4°C overnight for PCR clean-up the next day.

Cleaning up the PCR product

NOTE: Bring AMPure XP beads to room temperature (for 20 min).

CRITICAL STEP: Prepare fresh 80% (v/v) ethanol and 0.2 M NaOH.

23. To 22.5 µl of PCR product, add 11.25 µl of vortexed (30 s) AMPure XP beads and mix by pipetting (10 times).
24. Incubate at room temperature for 5 min.
25. Place the tube on a magnetic stand for 2 min.
26. While leaving the PCR tubes on the magnetic stand, carefully aspirate the supernatant.

    CRITICAL STEP: Do not aspirate beads. If aspirated, redo steps 25 and 26.

27. Add 100 µl of 80% ethanol and leave on the stand in the magnetic stand for 30 s.

    CRITICAL STEP: Do not resuspend the beads.

28. Aspirate out the supernatant carefully.
29. Add 100 µl of 80% ethanol and leave on the stand in the magnetic stand for 30 s.
30. Aspirate out the supernatant carefully.
31. Remove the tube from magnetic stand and allow to air dry in a tube stand for approximately 5 min.

    CRITICAL STEP: Visually check for cracks as over drying the beads will significantly reduce elution efficiency.

32. Add 26.15 µl of resuspension buffer and gently pipette 20 times to mix.
33. Incubate the tubes at room temperature for 2 min and then place on a magnetic stand for 2 min (until the supernatant cleared).
34. Transfer the supernatant (25 µl) to a new PCR tube.

    NOTE: The final supernatant can be stored at −15°C to −20°C for up to 1 week but we recommend proceeding to library normalization immediately.

Library normalization

35. Perform the Qubit DNA quantification method as described above to determine the genomic DNA concentration in cleaned up product.
36. Pool the genomic DNA from all of the tubes.

    NOTE: Sample with the lowest DNA concentration can be used in a volume of 10 µl as the reference to prepare a library pool using the formula: $\text{Volume}_{\text{required}}$ (V2) = $\text{Concentration}_{\text{original}}$ (S1) × $\text{Volume}_{\text{total}}$ (V1 = 10 µl) / $\text{Concentration}_{\text{required}}$ (S2).
37. To × µl of library pool, add × µl of freshly prepared 0.2 molar NaOH (final concentration 0.1 molar) and incubate for 5 min at room temperature.
38. To the NaOH treated suspension add an equal volume (2× µl) of LNS1. Label the tube as pooled amplified library (PAL).

    TIPS: In this modified step, normalize the library by using LNS1 (Library Normalisation Storage Buffer 1) only.
39. Dilute the PAL to 1:1000 by adding 1 µl of NaOH-LNS1 treated suspension to 999 µl of ultrapure distilled water.
40. Use KAPA library quantification kit (No ROX) to check the concentration of diluted pooled library in a real time PCR system using the following set up: a hot start run at 95°C for 10 min followed by 40 cycles of (95°C for 10 s and 60°C for 30 s).

    NOTE: Include a set of six DNA standards (with concentrations ranging from 20 pM to 0.0002 pM), three sets of negative control (ultrapure distilled water), and three sets of the DNA library in the qPCR run.
41. Determine the concentration of DNA in the pooled library by the standard curve method and calculate concentration in picomolar (pM) for each tube.

    NOTE: To calculate the original concentration of the pooled library we applied the formula:

    [Average sample concentration (in pM) × insert size standards (452 bp) × dilution factor (1000)] divisible by the Insert size of pooled library (500 BP)

    For example, for a qPCR determined concentration of 2.36 pM in a 1:1000 dilution of the pooled library, the library DNA concentration will be: (2.36 pM × 452 bp × 1000) / 500 bp = 2133.44 pM

    The value obtained from the calculation represents the concentration of DNA in the pooled library.
42. To estimate the dilution factor required to achieve a final library concentration of 15 pM in a 600 µl volume use the formula:

$$\begin{aligned}\text{Volume}_{\text{required}} &= (\text{Concentration}_{\text{required}} \times \text{Volume}_{\text{total}}) / \text{Concentration}_{\text{original}} \\ &= (15\ \text{pM} \times 600\ \mu\text{l}) / 2133.44\ \text{pM} \\ &= 4.22\ \mu\text{l}\end{aligned}$$

    NOTE: Therefore, 4.22 µl is added to 595.78 µl of HT buffer to produce a final concentration of 15 pM, in a final volume of 600 µl. The diluted library is then ready to be heat denatured and loaded into the MiSeq reagent cartridge.

Preparing pooled library for loading onto MiSeq

43. Thaw the PAL at room temperature and mix by pipetting up and down (5 times) followed by brief centrifugation.
44. Based on the library concentration example above, transfer 595.78 µl of HT buffer to a 1.5 ml diluted amplified library (DAL) tube containing 4.22 µl PAL.
45. Mix using a pipette (5 times).
46. Vortex the DAL tube at top speed, centrifuge briefly, and incubate exactly for 2 min at 96°C ± 2°C.
47. Immediately transfer the DAL tube to ice for at least 5 min or until loading.

CRITICAL STEP: Put the Illumina MiSeq sequencer through a short wash cycle to avoid cross-contamination of the DAL from previous usage.

48. Thaw the MiSeq reagent cartridge at room temperature [16].
49. Generate a MiSeq sample sheet using the Illumina Experiment Manager. See step 21 to identify primer sets for each sample.
50. Use the following configuration to set up the Miseq machine. Generate FASTQ workflow; FASTQ Only application; NexteraXT assay; 151 insert reads; assignment of the samples with a unique identifier and index-pair combination.

    CRITICAL STEP: Rinse the flow cell with MilliQ water and remove traces of water using soft tissue paper before inserting into the machine.

51. Transfer the entire 600 µl of DAL to the "Load here" well of the MiSeq reagent cartridge.
52. Following the setup procedure of the Illumina Experiment Manager, insert the cartridge into MiSeq instrument for sequencing to commence.

    TIPS: The raw FastQ sequence reads from whole-genome sequencing can be stored on the local computer as well as on the Illumina BaseSpace server (https://basespace.illumina.com/) for further analysis.

Bioinformatic analyses

NOTE: The selection of bioinformatics software for the analysis of WGS data will be determined by the objective of the study. Here, we used Geneious 9.1.8 (Biomatters Ltd.) [17], a desktop software to analyse our sequence data. Geneious was used to map the Fastq sequence reads to a publicly available reference genome for each species as follows:

53. Download Geneious from https://www.geneious.com/.
54. Go to File | Import | From File. Import raw-read files (Sample_xx_R1.fastq.gz and Sample_xx_R2.fastq.gz) into Geneious.
55. Download the Reference Genome from the NCBI database. For example, *Enterococcus faecium* NC_017960.
    - 55.1. In the Left panel | Go to NCBI | Nucleotide.
    - 55.2. Enter NC_017960 | Click Search.
    - 55.3. Once the genome has been found, click Download Full Sequence(s).
    - 55.4. Download the NC_017960 reference genome (The icon changes to a green circular genome when completed).
    - 55.5. Drag and drop the NC_017960 reference genome into the working folder.
56. Mapping the isolate sequence to the reference genome
    - 56.1. Hold CTRL and select both R1 and R2 raw read files (imported), and the reference genome (NC_017960) (downloaded).
    - 56.2. Click Align | Assemble|Map to Reference.
    - 56.3. Check the settings
        - 56.3.1. Reference Sequence = NC_017960
        - 56.3.2. Mapper = Bowtie2–fast and accurate read mapper
        - 56.3.3. Trim Before Mapping = Do not trim
        - 56.3.4. Results: Select all options
        - 56.3.5. Results | Save consensus sequences | Options
        - 56.3.6. Threshold = Highest Quality
        - 56.3.7. Threshold for sequences without quality = 95%
        - 56.3.8. No coverage call = '–'

56.4. When mapping to reference is complete, a new folder will be created containing four files:

56.4.1. Assembly Report

56.4.2. Consensus

56.4.3. Contig

56.4.4. Unused Reads

NOTE: Setting may vary depending on objective of analyses and quality of fastq reads.

NOTE: We also used open source databases, for example, TGS-TB [18], PhyResSe [19] and the Center for Genomic Epidemiology's [20] ResFinder and VirulenceFinder [21], to further analyse the whole genome sequence data of our selection of bacterial pathogens. These freely-available databases enable the acquisition of information on bacterial pathogens that included genotype and phylogeny, antibiotic-resistance mutations, and the presence of known virulence genes.

ANTICIPATED RESULTS

A consensus sequence was generated for each of the isolates analysed in Geneious. The Geneious report provided information on the percentage coverage of test sequence to the reference genome and the mean read depth (**Table 1**). Each contiguous sequence is viewable in Geneious and can be analysed for coverage with respect to the reference genome. Quality control checks of raw sequence data were also performed using FastQC [22]. This freely-available software provided information regarding per base sequence content and quality, per base and sequence GC content, and highlighted the parameters of the sequence quality.

Initial typing analysis

We used open source databases to analyze the sequence data. For example, Geneious mapped contiguous sequences were imported into PubMLST (https://pubmlst.org/) for sequence typing of *Haemophilus influenzae* and vancomycin-resistant *Enterococcus faecium*. This can also be achieved using raw fastq reads in the MLST profiling tool from the Center for Genomic Epidemiology (CGE) database (http://www.genomicepidemiology.org/). The Resfinder tool (https://cge.cbs.dtu.dk/services/ResFinder/) was used to identify acquired antimicrobial resistance genes from raw fastq files. For example, PubMLST typing classified NTHi 1 as sequence type 46 and Resfinder did not detect the presence of any antimicrobial resistance determining mutations.

Mycobacterium tuberculosis complex raw fastq.gz files were uploaded to the TGS-TB database (https://gph.niid.go.jp/tgs-tb/) to predict drug susceptibility, *in silico* spoligotype, lineage type, and phylogenetic classification. This database also enabled detection of IS6110 insertion sites, and 43 loci for variable number tandem repeat (VNTR) typing. The drug resistance profile of the MTBC isolates were further confirmed using PhyResSE database (http://phyresse.org/). For example, TGS-TB identified MTBC1 as a drug susceptible *Mycobacterium bovis* isolate.

Table 1. Percentage sequence coverage and mean read depth for each of the sequenced genomes with respect to reference strains.

Sample	Reference sequence coverage (%)	Mean read depth
VRE1	90.98	197.4
VRE2	89.4	178.4
VRE3	90.96	168
NtHi1	93.8	263.5
NtHi2	86.5	61.3
NtHi3	89.1	124.9
MTBC1	96.4	104.4
MTBC2	96.8	100.4
MTBC3	96.8	68.4

Coverage refers to the percentage of reference genome bases covered by mapped sequence reads. Mean read depth indicates the mean number of times each base is mapped by a sequence read. Reference genomes used were *E. faecium* ST18 DO (TX16) (accession number NC_017960), *Haemophilus influenzae* 86-028NP (nontypeable) (accession number NC_007146), and *Mycobacterium tuberculosis* H37Rv (accession number NC000962). VRE, vancomycin resistant *Enterococcus faecium*; NtHi, non-typeable Haemophilus influenza; MTBC, *Mycobacterium tuberculosis* complex.

TROUBLESHOOTING

Possible problems and their troubleshooting solutions are listed in **Table 2**. There are a number of limitations associated with the protocol that should be noted. These include: effective results with the protocol are reliant on the efficacy of the extraction procedure in producing a

sufficient quantity of genomic DNA; analysis of sequences generated on an Illumina platform can be affected by the presence of highly repetitive regions; and depending on the output information sought, genome assembly can be influenced by the reference genome selected for the mapping of reads. Nevertheless, the protocol was effective in generating high quality sequencing data for the range of bacterial species tested.

Table 2. Troubleshooting table.

Step #	Problems	Causes	Suggestions
23	Low concentration of AMPure XP bead captured purified products	Bead clean-up affects the quality and quantity of amplified libraries that will be included in downstream sequencing process	Make sure AMPure XP beads are held at room temperature for 20 min before starting the clean-up process. Furthermore, ensure that 80% ethanol is freshly prepared
35	Variation in the concentration of amplified library	The concentration of input DNA used for library preparation affects the final yield of genomic data	Measure the concentration of input DNA using Qubit fluorometer rather than a nanodrop and make the appropriate dilution for a DNA concentration of 0.2 ng/µl
56	Poor sequencing results	The choice of forward and reverse index primer set affects the sequencing of libraries prepared	Avoid primer combinations S502 with N705 /706, and S503 with N701/702
56	Poor sequencing results	Effective denaturation of pooled library not achieved after bead clean-up	Ensure that the NaOH is freshly prepared at the correct concentration
56	Poor sequencing results	Repeated thawing and freezing of the pooled library reduces the quality of sequence reads generated	Before preparing the pooled library for loading onto MiSeq, ensure that the machine has already been appropriately cleaned after the previous run and has sufficient storage space (at least 25 GB)

Acknowledgments

This research was supported by funding from the Royal Hobart Hospital Research Foundation (17-104) and the Tasmanian Community Fund (36Medium00014).

References

1. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, et al. (2015) Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 15: 1193-1202. doi: 10.1016/S1473-3099(15)00062-6. PMID: 26116186
2. Takiff HE, Feo O (2015) Clinical value of whole-genome sequencing of Mycobacterium tuberculosis. Lancet Infect Dis 15: 1077-1090. doi: 10.1016/S1473-3099(15)00071-7. PMID: 26277037
3. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, et al. (2012) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis 13: 130-136. doi: 10.1016/S1473-3099(12)70268-2. PMID: 23158674
4. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al. (2012) Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci USA 109: 3065-3070. doi: 10.1073/pnas.1121491109. PMID: 22315421
5. Llarena A, Taboada E, Rossi M (2017) Whole-Genome Sequencing in Epidemiology of Campylobacter jejuni Infections. J Clin Microbiol 55: 1269-1275. doi: 10.1128/JCM.00017-17. PMID: 28249998
6. Reuter S, Harrison TG, Köser CU, Ellington MJ, Smith GP, et al. (2013) A pilot study of rapid whole-genome sequencing for the investigation of a Legionella outbreak. BMJ Open 3: e002175. doi: 10.1136/bmjopen-2012-002175. PMID: 23306006
7. Gautam SS, Mac Aogáin M, Cooley LA, Haug G, Fyfe JA, et al. (2018) Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. PLoS One 13: e0192351. doi: 10.1371/journal.pone.0192351. PMID: 29466411
8. McGinnis J, Laplante J, Shudt M, George KS (2016) Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J Clin Virol 79: 44-50. doi: 10.1016/j.jcv.2016.03.005. PMID: 27085509
9. Park DJ, Dudas G, Wohl S, Goba A, Whitmer SLM, et al. (2015) Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. Cell 161: 1516-1526. doi: 10.1016/j.cell.2015.06.007. PMID: 26091036
10. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, et al. (2017) Zika virus evolution and spread in the Americas. Nature 546: 411-415. doi: 10.1038/nature22402. PMID: 28538734
11. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, et al. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 366: 2267-2275. doi: 10.1056/NEJMoa1109910. PMID: 22693998
12. Halachev MR, Chan JZ, Constantinidou CI, Cumley N, Bradley C, et al. (2014) Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England. Genome Med 6: 70. doi: 10.1186/s13073-014-0070-x. PMID: 25414729
13. Davis RJ, Jensen SO, Van Hal S, Espedido B, Gordon A, et al. (2015) Whole Genome Sequencing in Real-Time Investigation and Management of a Pseudomonas aeruginosa Outbreak on a Neonatal Intensive Care Unit. Infect Control Hosp Epidemiol 36: 1058-1064. doi: 10.1017/ice.2015.133. PMID: 26050101
14. Leong KWC, Cooley LA, Anderson TL, Gautam SS, McEwan B, et al. (2018) Emergence of Vancomycin-Resistant Enterococcus faecium at an Australian Hospital: A Whole Genome Sequencing Analysis. Sci Rep 8: 6274. doi: 10.1038/s41598-018-24614-6. PMID: 29674657
15. Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 10: 21-8. doi: 10.1186/1746-4811-10-21. PMID: 25053969
16. Assad NA, Balmes J, Mehta S, Cheema U, Sood A (2015) Chronic obstructive pulmonary disease secondary to household air pollution. Semin Respir Crit Care Med 36: 408-421. doi: 10.1055/s-0035-1554846. PMID: 26024348
17. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649. doi: 10.1093/bioinformatics/bts199. PMID: 22543367

18. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, et al. (2015) TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. PLoS One 10: 2015-10. doi: 10.1371/journal.pone.0142951. PMID: 26565975

19. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, et al. (2015) PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol 53: 1908-1914. doi: 10.1128/JCM.00025-15. PMID: 25854485

20. Thomsen MCF, Ahrenfeldt J, Cisneros JLB, Jurtz V, Larsen MV, et al. (2016) A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome Sequencing Data for Clinical Diagnostics and Surveillance. PLoS One 11: e0157718. doi: 10.1371/journal.pone.0157718. PMID: 27327771

21. Kleinheinz KA, Joensen KG, Larsen MV (2014) Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4: e27943. doi: 10.4161/bact.27943. PMID: 24575358

22. Babraham Bioinformatics (2010) FastQC: a quality control tool for high throughput sequence data. [cited October 4, 2018]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Response to the Examiner's Comments

Examiner-1 comments and responses

Strengths: There are several interesting and significant findings described in this thesis.

Identification of a set of genes absent in H37Rv but present in the Rangipo strains (in particular the five genes shown in Figure 3.1) is a major finding that can help explain differences between the Rangipo strains and H37Rv. The genes that are present only in the Rangipo stains appear to be associated with vital cellular functions. This may mean that cellular physiology of Rangipo strains is not identical to that of H37Rv. In such a case, the transmissibility or virulence of Rangipo strains could be influenced by these genes. For these reasons, the reported findings are significant.

Molecular epidemiological studies described in Chapter 4 of M. tuberculosis strains isolated in Tasmanai in 2014-2017 are straightforward, thorough and conclusions made should be useful in making relevant public health policy decisions. Equally interesting is the TASMDR1 case. Seeking out VTB1 to undertake a detailed analysis of TASMDR1 was a very good strategy.

The current draft of this thesis should be revised to address the minor comments, queries and recommendations that I have made. Most of these comments are minor, but some require clarification of the statements made or methods used.

Response: I thank the examiner for acknowledging the significance of the work undertaken. The grammatical errors noted by the examiner have been corrected in the revised version of the thesis. The thesis has been thoroughly proof read before final submission to the chair of examiners. Here, I include responses to the comments, queries and recommendations made by examiner 1.

To be precise, this is 'resequencing' as the reference genome was used to generate 'a complete' genome sequence for each strain.

Response: Thank you for the comment. In this work, new, previously-unsequenced isolates of *M. tuberculosis* were sequenced. In order to scaffold the FASTQ sequence reads, an established reference genome of *M. tuberculosis* was used i.e. H37Rv. Therefore, re-sequencing of previously-sequenced isolates of M. tuberculosis was not performed in this work. To fully clarify this point, the following statement has been added to the revised thesis: Whole-genome sequencing of nine isolates representing the Rangipo genotype was performed on an Illumina MiSeq platform. The sequence data of each isolate was mapped to the reference *M. tuberculosis* strain H37Rv to generate a consensus genome. Page ii.

1. This statement is not necessary.

   Response: The statement, 'WGS offers the capability to decipher the biology and perform molecular epidemiological typing of *M. tuberculosis*' has been removed as suggested by the examiner. Page v.

2. To be factually correct, I recommend adding 'sputum smear microscopy based on AFB staining' as this modality is still extensively used in resource limited setting (where most of the TB patients are first diagnosed and treated).

   Response**:** The recommended information has been added to the revised thesis with the following statement: The first group experience symptoms of TB (for example fever, weight loss and haemoptysis), transmit infection and are usually detected by sputum smear microscopy and culture-based diagnostic tests. Secondly, patients with latent TB, although infected, do not exhibit symptoms or transmit the infection and are undetectable in culture-based tests (Barry et al., 2009; Desikan, 2013; Esmail et.al, 2014). Page 2.

3. No WGS is not suited for predicting functional roles of genes. WGS is good at providing raw DNA sequence at high resolution and accuracy. 'Functional roles' cannot be predicted with sufficient accuracy as WGS is not meant for this purpose.

   Response: The following statement' Whole-genome sequencing poses an advantage to ideally identify all the clinically relevant mutations and predict

wide range of drug resistance determining mutations (Wells et al., 2013) as compared to Xpert MTB/RIF and line probe assays which are limited to detect mutations in specific gene locus (Sanchez-Padilla et al., 2015).' has been added in the revised thesis after accepting the examiners comment. Page 16.

4. What aspects of the data were analysed? How were they validated? Briefly describe.

   Response: The whole genome sequence data obtained as Fastq files were analysed for variant detection, infer phylogenetic relationship, establish transmission networks, and identify differential genome content. Data analyses for specific experiments have been described in the methods section of each chapter. A new table (table 3) has been added in the revised thesis where tools and databases used for analyses of genomic data in this study have been summarised. We did not validate any bioinformatics pipeline during the study which is one of the limitations of the study and is subject to future study. Page 55.

5. Very important finding. Also, please verify if there is any sequence homology of Rv3109 and Rv0869c to the genes that you have identified in the Rangipo strain (*MoaA* and *MoaB* in the Rangipo strain as shown in Figure 3.1).

   Response: Thank you for recognizing our work. A preliminary investigation of genome sequence data of Rangipo strain suggests the presence of sequence homology of Rv3109 and Rv0869c with 100% similarity at the nucleotide level in the Rangipo strain. Page 73.

6. A more stringent assessment would have been Southern blot analysis. You should mention that Southern blotting was not performed.

   Response: To clarify the method used in confirming the presence of differential genes in Rangipo strains following sentence has been added in the updated thesis: PCR rather than Southern Blotting was performed on the Rangipo isolates and H37Rv using primers for the genes (MT3426,

MT3427, MT3428, MT2082 and MT1360) to confirm their presence and absence as indicated by the genome sequence analysis. Page 72.

7. Which mutations were considered?

   Response: Drug resistance determining mutations were predicted using PhyResSe database (Feuerriegel et al., 2015) which identifies the mutations related to first- and second-line drug resistance in tuberculosis. A new statement has been added to the main text including examples of some of the mutations detected by PhyResSe. Page 74.

8. You might also want to state (if you have the data) whether there are genes in H37Rv that are not present in the Rangipo strains.

   Response: Thank you for the comment. A preliminary analysis based on whole genome sequence data showed the absence of four genes (with functional annotation and excluding PE, PPE, PGRS, transposes and prophages) in the Rangipo strain that were present in the reference H37Rv. These include, Rv1758 (cutinase 1), Rv1760 (diacylglycerol acetyltransferase), Rv2653c (toxin), Rv2654c (antitoxin). This has been added to the conclusion and recommendation section 3.4 of the updated thesis. Page 81.

9. While using H37Rv genome as template for assessment of known mutations that confer resistance to the TB treatment drugs (relevant in the context of this specific clinical isolate) is fine, it appears a bit strange that you chose not to use other more clinically relevant strains (such as CDC1551, Erdman etc) especially since you recommend using multiple stains for alignment studies (in the previous chapter).

   Response: This particular analysis (section 5.3.2) was done to deposit the genome sequence data of TASMDR1 in the public databases. Here, we did not aim to compare the genome of TASMDR1 with different reference strains and therefore it was mapped against H37Rv to identify SNPs and drug resistant determining mutations. Page 105.

Examiner-2 comments and responses

Thank you for asking me to be an external examiner for this thesis by Sanjay Gautam for consideration for the degree of Doctor of Philosophy. Overall I feel that it meets the requirements to be awarded a PhD but there are some concerns and minor points that I detail below that I do feel need to be addressed.

Response: I would like to thank the examiner for acknowledging and reviewing the work undertaken. The grammatical errors noted by the examiner have been corrected in the revised version of the thesis. The thesis has been thoroughly proof read before final submission to the chair of examiners. Here, I include responses to the comments, queries and recommendations made by examiner 2.

1. There is also a constant flux between British and US English throughout the thesis. I recommend choosing while style and then being consistent.

   Response: The discrepancies between British and US English have been corrected by adopting the British English format.

2. Finally, there are issues in the latter part of the thesis with distinctions between protein and gene names. Genes start with a lower - case letter and are italicised, proteins start with a capital letter and are not italicised. There are many references to italicised capitalised names and it is not clear if one is referring to the protein or the gene.

   Response: The gene names have now been corrected to lower case and italicised while the name of the protein has not been italicised and start with the capital letter.

3. Much of the content of the thesis appears to be published. It was not clear why the published papers were simply not included as chapters as presumably these would have been freer of errors and already peer - reviewed. Please ignore this comment if the University of Tasmania does not allow for mixed or theses by publication.

Response: The thesis format used is one of the accepted formats for PhD theses at the University of Tasmania whereby published and under review papers are appended but a full thesis with conventional chapters is also written i.e. Page 5 of the Higher Degree by Research Thesis Preparation, Submission and Examination Policy, University of Tasmania “2. a written text containing a combination of peer reviewed publications, articles undergoing peer-review or revision, and conventional chapters presented as typescript”.

4. I note that some of the content is in press or has been submitted. Could updates be provided as to whether this has now been published.

   Response: The paper entitled “Potential cross-border translocation of multi-drug resistant tuberculosis in latent form to Australia” and A Step by-Step Beginner’s Protocol for Whole-genome Sequencing of Human Bacterial Pathogens” is now published while next paper “.

5. Chapter 2 is a methods paper – I will provide some more details below but the changes to well-known and used protocols that are suggested do not appear to be validated. There is no comparison made between the manufacturers’ protocols and the submitted protocol and therefore there is an outstanding question around the validity of the changes made.

   Response: The submitted protocol was based primarily on the manufacturer’s protocol except for the consumables used and specific reaction steps. We modified a number of the DNA extraction steps to obtain a sufficient quantity of contamination free template. Similarly, we replaced library normalization plates and Nextera XT tagment amplicon (NTA) plates with conventional polymerase chain reaction (PCR) PCR tubes which may represent a cost-effective alternative. In addition, we have recommended the use of equal DNA concentrations of each library during library normalization to ensure better coverage and minimize bias. The protocol was adopted based on its reproducibility and reliability and has been cited across many of our publications (Mac Aogain et al, 2016, Gautam et al., 2017, Gautam et al., 2018, Leong et al., 2018).

6. The bioinformatics analyses presented throughout the thesis are confusing. They often switch between multiple pieces of software used for very different purposes i.e. mapping tools, then de novo assembly, then calling SNPs, etc. The purpose of a methods section is to be detailed enough to allow one to reproduce the analyses. While the wet lab sections do this well, I do not believe anyone could reproduce the findings of the bioinformatics analysis as simple things like parameters are not given.

   Response: The switches between the data analyses tools and databases was due to the objective of the study. For example, SNPs were called either to identify drug resistance determining polymorphisms or to establish transmission network. The data analyses pipeline differed in each of these cases. The bioinformatic analyses steps outlined in chapter 2 includes basic steps in genome assembly and SNP detection. The objective specific use of different pipelines has been discussed within the methods section of each chapter. To make the data analyses pipeline clearer, a new table (Table 3) has been added in chapter 2 of the revised thesis where database and tools utilized have been discussed.

7. There are multiple references in the thesis to tuberculosis being caused by a single infectious agent, but then the acknowledgement elsewhere that it is caused by a complex of species. Please be consistent – the disease of tuberculosis is caused by a number of species that make up the Mycobacterium tuberculosis complex.

   Response: Thank you for specifying inconsistencies. Relevant changes have been made. For example, a revised statement ‘Tuberculosis (TB) is a major cause of global mortalities causing 1.6 million deaths in 2017 alone due to different species of *Mycobacterium tuberculosis* complex’ has been included in the abstract. Page ii.

8. “the colon tissue biopsy specimen detected the presence” – a specimen cannot detect things in itself. Therefore specific what testing was performed on the specimen to identify the bacteria.

Response: Mycob*acterium tuberculosis* were detected in the colon tissue biopsy sample using MGIT culture system. The suggested edits have been accepted in the revised draft of the thesis. Page iv.

9. Brosch et al. 2002 is a very old reference when talking about TB ancestry and there is much more newer data available. I would recommend https://www.nature.com/articles/nrmicro.2018.8 instead. M. africanum is less likely to be referred to as a separate species now – make sure to read Gagneux's latest papers on this. Page 1.

   Response: I thank the reviewer for the suggestion. The recommended reference, Gagneux, 2018 has been added in the revised thesis.

10. "BCG vaccine was covered in more than 86%" – use of language. Does this mean it is on the essential medicines list or national immunisation programmes of over 86% of countries?

    Response: The statement was made to represent following data, "In 2016, the global BCG vaccine coverage estimate was 90% reported from 169 countries, and all WHO regions have an average BCG vaccine coverage greater than 86%". The following statement has been included in the revised thesis: The preventive Bacillus Calmette –Guérin (BCG) vaccines average coverage was more than 86% across all WHO regions (World Health Organization, 2017b). Page 5.

11. Note that there are errors in Zhang et al. 2013 analysis of BCG. The more updated ref is https://www.nature.com/articles/srep15443

    Response: The Zhang et al. 2013 paper has been replaced with the Abdallah et al. 2015 paper. Page 7.

12. BCG is given routinely in London, and this doesn't meet the high-risk definitions given in the text

    Response: We agree with the examiner's comment. "Europe" has been updated with Western European countries as mentioned in the original text. Page 8.

13. Some of the discuss around membership of the M. tuberculosis complex is a little outdated. Current thinking is that they are variants of the same species: https://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.002507#tab2

    Response: Thank you for the reference. A new statement from the recommended reference (Riojas et al. 2017) has been incorporated in the text. Page 19.

14. it is wrong to say that TB treatment is the longest regimen to treat an infectious agent. There are bacterial diseases that need longer courses of treatment, e.g. chronic Lyme and Q fever, and viral diseases that need lifelong therapy, e.g. HIV

    Response: This statement has been modified to: “Treatment of TB, therefore, involves one of the longest regimes directed against an infectious disease” Page 26.

15. The findings from Gardy et al. are misquoted here – recommend looking more closely at the findings described in the paper and the impact of MIRU versus SNP-based typing.

    Response: Thank you for the comment. The sentence has been updated with: Whole-genome sequencing identified identical MIRU-VNTR genotypes which along with the analyses of patient’s social networks detected the presence of a common progenitor during a TB outbreak that existed before the onset of transmission (Gardy et al., 2011). This outbreak was further related to a social event of the use of crack cocaine (Gardy et al., 2011). Page 38.

16. As noted above, it is all well and good to describe a new wet lab protocol but there are no results demonstrating that the results from it are as good as published protocols.

    Response: The aim of introducing modifications in the manufacturer’s protocol was to generate a step-by-step protocol that can be used by

laboratory personnel with no prior experience with next-generation sequencing. There was no comparison made with any published protocols rather we aimed to reproduce the method to obtain high quality genomic data for downstream analyses. This method of generating whole-genome sequence data of bacteria including *M. tuberculosis* was based on the manufacturer's instructions (Illumina, Inc. 2012). Modifications were done in terms of the usage of consumables and library normalization step to maximize the quality of output data (Gautam et al., 2019a). The proposed method was not compared for its sensitivity and accuracy with any published protocol; however, we were able to reproduce the protocol for studies involving *M. tuberculosis* and other bacterial pathogens (Gautam et al., 2019; Gautam et al., 2018; Gautam, Mac Aogáin, & O'Toole, 2017; Leong et al., 2018; Mac Aogáin et al., 2016). Page 56.

17. Also as noted above, there is lots of confusion when it comes to describing the bioinformatics pathways. For example, FastQC is mentioned twice, as a standalone piece of software and as part of Galaxy. Geneious is used to map and find variants but then BWA and SAMtools are mentioned. No details are provided around parameters for any software used. The contribution by the collaborator at Trinity College Dublin is also not clear. How has the final pipeline been validated, particularly against published pipelines?

    Response: The repeated information regarding FastQC has been deleted in the revised thesis. Geneious was initially used for the mapping and variant calling. Later, BWA and SAMtools were used by a collaborator at Trinity college Dublin to perform similar analyses. In this way, we used different platforms to reach a final conclusion. No validation experiments were performed for a new pipeline, rather we based our analyses on the published works of literature to design cut-offs and establish a working

pipeline. This was one of the limitations of the bioinformatic approach, but this does not significantly detract the epidemiological analyses done in the current study. Page 55.

18. I would strongly recommend against using random multiple base matching – why would you only map uniquely and exclude reads that could match multiple places. Otherwise, SNP information is extremely difficult to interpret. Also throughout this and subsequent chapters here is mention of SLVs and SNPs, but no description of indels that can be detected and the sizes obtained. In addition there is constant confusion in the methods sections about mapping versus de novo assembly, why you would use one over the other, etc.

    Response: We employed random multiple base matching as mapping and SNP detection using BWA and SAMtools allowed us to detect SNPs and genes that were completely missing in the Rangipo strain consistently. The presence/absence of genes were confirmed by PCR amplification and Sanger sequencing. The five genes that were detected in the Rangipo strain but absent in H37Rv were regions of insertions in ORFs in Rangipo strain relative to H37Rv. Similar insertion events were reported by Fleischmann et al.,2002 in strain CDC1551. In this study, we were more focused on detecting differential gene content in the Rangipo strain and therefore we did not report all the indels. Furthermore, we performed de novo assembly only when we were looking to deposit the genome in the public database. All other analyses were done after mapping the raw sequence with different reference genomes. Page 63.

19. Lots of details missing about the phylogenetic methods used, particularly choice of GTR substitution model, methods for generating trees, bootstrapping, etc. There is also a lack of information about how displayed trees in this thesis were rooted.

    Response: For phylogenetic tree building, PhyML was used where 10 starting trees were selected and bootstrapping was performed over 100 replicates. Furthermore, the tree was rooted to *Mycobacterium canetti*. This information has been added in the revised thesis. Page 73.

20. The section on consensus genome assembly is slightly out of place. Why was this done on only one isolate? Why not assemble all of them? How was the assembly improved beyond a simple contig assembly?

    Response: Prior to this work, there were no publicly available genome sequences of Rangipo strains of *M. tuberculosis* in the DDBJ/ENA/GenBank databases. The purpose of section 3.2.4.5 was to generate and publish an assembled consensus genome of *M. tuberculosis* Rangipo. Strain 494 was selected as a representative genome of the Rangipo strains to enter into DDBJ/ENA/GenBank. The genomes of the other eight Rangipo isolates differed by ≤15 SNPs from strain 494. Page 74.

21. It is not clear what conclusions around virulence you are actually trying to reach. The mutations in virulence genes would appear to support reduced virulence for Rangipo strains, but you overall conclude the opposite. Is this simply due to inclusion of similar genes as CDC1551?

    Response: Here we reported polymorphisms in genes that have previously been related to the virulence of *M. tuberculosis* in the published literature. For example, disruption of *aceAa* has been associated with the reduced survival of *M. tuberculosis* in the lungs and spleen of infected mice (Munoz-Elias and McKinney JD, 2005). To determine the effect of the polymorphisms identified in the Rangipo strain on the virulence of *M. tuberculosis*, downstream functional studies involving site-directed mutants would need to be conducted. Page 76.

22. There are two major weaknesses in this chapter. The first is the selection of strains – why were a single year's group of strains not sequenced. There were not many so it wouldn't be too expensive.

    Response: As Tasmania is a low TB incidence state (approx. 1.6/100,000), the number of TB cases from a single year is very low. Furthermore, not all *M. tuberculosis* specimens were culturable. In addition, there is no single reference laboratory for TB in Tasmania which meant that isolates had to be sourced from multiple intra- and inter-state laboratories. To reach the

number of isolates examined in this thesis work, it was necessary to examine isolates from a greater number of calendar years.

23. Second, there is an important finding of BCGosis that is completely ignored in the results and discussion. The only real demonstration is in TABLE 6. It has been described as tuberculosis, which is incorrect, but is this simply a result of bladder cancer treatment with BCG?

    Response: The BCG was isolated from urine specimen taken from a patient undergoing bladder cancer treatment with BCG. It was included in the study to provide a genetic comparison with the *M. bovis* TB case in order to establish that the latter TB case was indeed caused by *M. bovis* and not BCG. A new paragraph has been added in the revised thesis regarding the preliminary analysis of the BCG strain. Page 94.

24. How were the 18 strains chosen? Why not sequence all? What is the risk of bias introduced here?

    Response: There were 29 TB cases notified in Tasmania during the calendar years, 2014-2016. All of the bacteriologically confirmed cases with isolates available (n=18) were included in this study for whole genome sequencing. Page 84.

25. You describe a strange mix of pipelines in chapter 3 so which was actually used for this chapter?

    Response: To make it clearer, section 4.2.4 has now been updated and chapter 4 specific data analyses pipeline included. Page 85.

26. “Osteomyelitis” isn’t a specimen. Was it pus or bone? Similarly “tissue” is not enough description.

    Response: The information presented was based on the available public health surveillance records. The available record of the patient did not explain if it’s a pus or bone specimen. The tissue sample was obtained from a colon biopsy and this has been revised in table 7. Page 88.

27. M. bovis BCG detection in urine is not tuberculosis.

Response: I agree with the examiner's comment. The BCG was isolated from a urine specimen taken from a patient undergoing bladder cancer treatment with BCG. It was included in the study to provide a genetic comparison with the *M. bovis* TB case in order to establish that the latter TB case was indeed caused by *M. bovis* and not BCG. Page 88.

28. Similar problem with phylogenetics. Methods? Bootstraps? Root?

    Response: For phylogenetic tree building, PhyML was used where 10 starting trees were selected and bootstrapping was performed over 100 replicates. Furthermore, the tree was rooted to *Mycobacterium canetti*. This information has been added in the revised thesis. Page 90.

29. "These cases of drug susceptible pulmonary TB were detected in the gastric aspirate" how is it pulmonary if in a gastric sample?

    Response: The gastric aspirate samples were obtained for pulmonary TB detection in infants. We confirmed these as cases of pulmonary TB in consultation with the treating physician. Page 88.

30. I would expect an interesting finding such as *M. bovis* would be described more fully. Age of patient? Symptoms? Presentation, etc?

    Response: The clinical data that was available for this case was limited. The patient variables available consisted of age, specimen type, and comorbidity (section 4.3.4) and has been reproduced in section 4.4 of the updated thesis. Page 97.

31. The methods section only focuses on sequencing a single isolate, yet the results talk about a second isolate being sequenced.

    Response: The same method was applied for sequencing both the isolates. The first isolate TASMDR1 was sequenced during our work described in chapter 4. The second sample, VTB1, was obtained later following travel to Viet Nam to collect the isolate from Ho Chi Minh City and then sequenced.

32. Aforementioned confusion between mapping and de novo assembly.

Response: Thank you for the comment. Mapping with the reference strain H37Rv was done primarily to identify the number and types of single nucleotide polymorphism and detect the presence of drug resistance determining mutation. *de novo* assembly was done to assemble the contigs so that they can be ordered with respect to the reference genome for their deposition in the public database (Gautam et al.,2018). Page 102.

33. I am really unclear what the final sentence is saying here

    Response: The final sentence in page 103, “The number of variants in the MDR-TB strain with respect to the H37Rv reference genome were determined using the SAM-tools analyses suite” is describing the process of determining the number of variants present in the MDR-TB strain as compared to *M. tuberculosis* reference strain, H37Rv. Furthermore, we incorporated the patient information along with the SNP analyses data to infer a transmission network. Page 103.

34. Are the resistance determining mutations similar or identical to TASMDR1?

    Response: The resistance determining mutations present in TASMDR1 and VTB1 were exactly the same. The sentence has been rewritten as, ‘The drug resistance profile and lineage typing of VTB1 was identical to TASMDR1, i.e. VTB1 belonged to East Asian L2 (Beijing) and conferred exactly the same drug resistance determining mutations as of TASMDR1 i.e. isoniazid, rifampicin, pyrazinamide, ethambutol and streptomycin.’ Page 108.

35. No evidence is presented about improving the methodology. What was the baseline protocol used in Dublin, how was it improved and how was the improvement validated?

    Response: Thank you for the comment. The protocol was originally designed at Trinity College, Dublin and was later adopted at the University of Tasmania. The new thing that we did in this protocol was the replacement of consumables and certain steps recommended by the manufacturer. The major advantage was decrease in the cost related to consumables. Furthermore, this protocol worked well for Gram positive, Gram negative

and acid-fast bacilli. Some of the changes in protocol across the experiments include the process of obtaining bacterial DNA from the liquid culture which is dependent upon cell wall composition of the bacteria. Based on this protocol, we have performed experiments on Gram positive, Gram negative and acid-fast bacilli and have been able to reproduce each time with the generation of high -quality data (minimum read depth of 20). Contrarily, we did not validate any bioinformatics pipeline during the study which is one of the limitations of the study and is subject to future study. Page 115.

36. Linnaean names are not properly italicised in titles of papers

    Response: Thank you for the comment. The suggested edits have been made.

37. Reference half missing

    Response: This incorrect reference has now been deleted. Page 167

    Response to in-text comments by examiner 2 not listed in the comment file.

38. If state this then it is worth then stating what the proportions are in HIV - infected and uninfected patients.

    Response: The data on the proportion of HIV-infected and uninfected patients have been updated in the revised thesis and include the following statement, 87% (n=1255) of TB cases in Australia were detected among those with known HIV status where 27 cases were associated with HIV positive individuals. Sixty two percent had pulmonary TB of which 87% were bacteriologically confirmed (World Health Organization, 2018). There were 25 pulmonary TB cases identified as MDR/RR-TB during 2017 (World Health Organization, 2018). Page 6.

39. Mono or MDR?

    Response: These were categorised as MDR/RR-TB. The statement has been updated. Page 6.

40. Has chapter 3 been published?

    Response: Yes, Chapter 3 is published in the Journal of Infectious Diseases and can be accessed here: https://doi.org/10.1080/23744235.2017.1330553. Page 62.

41. What about Indels? Particularly for PZA resistance?

    Response: Yes, indels were determined for PZA resistance using PhyResSe database and sequence analyses done after BWA/SAMtools mapping and variant finding. Page 65.

42. Was this tested and found to be the best model (e.g. using Akaike or Bayesian information criteria?)

    Response: Unfortunately, we did not aim to test different models of phylogenetic tree building. Page 64.

43. Filtering criteria for SAM-tools?

    Response: The SAMtools Mpileup was run with the following command: SAMtools Mpileup -DSugBf $ref -Q30 -q30 -o40 -e20 -h100 -m2
    A SNPSIFT quality filter of 'QUAL >= 999' was imposed on detected variants with minimum 5-fold coverage across all strains analyzed. Page 65.

44. But also, as you note, a number of lineage 4 strains.

    Response: Yes, these genes (some but not all) were also found in other Lineage 4 strains of *M. tuberculosis*. In this section we were more focused on the outbreak strain. Page 76.

45. Details of the BCG case? Type of BCG? Was it mapped to the genomics data available on BCG?

    Response: This case of *M. bovis* BCG was related to a patient obtaining BCG for his bladder cancer treatment and was therefore not investigated further. However, an analysis of the genome was performed using CGE database (http://www.genomicepidemiology.org/). A paragraph regarding Tasmanian BCG strain has been added in the revised version of the thesis. Page 94.

46. Size range of insertions/deletions in TASMDR1

Response: The indels in TASMDR1 isolate ranged in their size of 1 to 501 base pairs. Page 105.

47. This wasn't clear - what were the new steps? What are the benefits?

Response: Thank you for the comment. The protocol was originally designed at Trinity College, Dublin and was later adopted in University of Tasmania. The new thing that we did in this protocol was the replacement of consumables and certain steps recommended by the manufacturer. The major advantage was decrease in the cost related to consumables. Furthermore, this protocol worked well for Gram positive, Gram negative and acid-fast bacilli. Some of the changes in protocol across the experiments include the process of obtaining bacterial DNA from the liquid culture which is dependent upon cell wall composition of the bacteria. Based on this protocol we have performed experiments on Gram positive, Gram negative and acid-fast bacilli and have been able to reproduce each time with the generation of high -quality data (minimum read depth of 20). Contrarily, we did not validate any bioinformatics pipeline during the study which is one of the limitations of the study and is subject to future study. Page 115.